to appear
Proc. RAGSQUAD, Berkeley

On the trace formula for quadratic forms and some applications

by

Eberhard Becker and Thorsten Wörmann

Mathematisches Institut der Universität Dortmund
4600 Dortmund, Fed. Rep. of Germany

Introduction

This paper deals with a variant of the trace formula for quadratic forms which allows applications to some algorithmic problems of real algebraic geometry. The formula will be applied to the counting of real zeros on 0-dimensional varieties under side constraints, as well as to the 0-dimensional case of the Bröcker-Scheiderer result about the description of basic open semi-algebraic sets. Furthermore, it can be used to give a 'visible' argument for Tarski's theorem on Quantifier Elimination in the theory of real closed fields.

It is only fair to admit that our method is nothing but a modern version of old ideas of Hermite-Sylvester who had already shown how to count real zeros by calculating signatures of appropriate quadratic forms. What has been added to their approach is a certain algebraic machinery that enables us to treat multivariate problems more uniformly. In an analogous approach P. Pedersen has independently developed a similiar method to count real zeros, also starting with Hermite ideas.

1 The trace formula

Trace formulae of various kinds frequently occur in the literature on quadratic forms. In particular, M. Knebusch made an extensive study of such formulae, cf. [K1], [K2], [K3]. So, our topic is basically well-known. However, in our approach we will formulate a trace formula under much weaker assumptions which allows new applications to real algebraic geometry. The main difference is that degenerate forms are allowed. In addition, the ring extensions involved need not be Frobenius and, moreover, as in Mahé's work [M, (4.11)], the real spectrum of rings instead of the space of signatures is used.

Our framework is the following. Let A be any commutative ring, $\tau : A \to B$ any ring homomorphism rendering B into a finitely generated projective A-module. Moreover, let M be a finitely generated projective B-module, equipped with a B-bilinear map $\varphi : M \times M \to B$. Note that the non-degeneracy of φ is not assumed. If $s : B \to A$ is any A-linear map then we can define the transfer $s_*(\varphi)$ which is an A-bilinear map on M, considered as an A-module, and which is defined as follows:

$$s_*(\varphi) : M \times M \to A, (m,n) \mapsto s(\varphi(m,n))$$

We will primarily deal with the trace map

$$s = tr_{B/A} : B \to A,$$

cf. [DeM, I; III, §2]. Nevertheless, other linear forms occur as well and deserve a study in their own right as will be indicated at the end of this section.

The ring homomorphism $\tau : A \to B$ induces a 'restriction' map $\tau^* : Sper\, B \to Sper\, A$ between the real spectra of A and B. We will freely use notions and results from the theory of the real spectrum as described in [C-R], [BCR], [Bel]. Note that $Sper\, A = Spec_r A = R - Spec\, A$ etc. if other notations are used. If $\alpha \in Sper\, A$ is given, any $\beta \in Sper\, B$ with $\tau^*(\beta) = \alpha$ is called an extension of α, written: $\beta \mid \alpha$. By [C-R, 4.3, p. 40] or [Bel] one knows that the canonical map $B \to B \otimes_A k(\alpha)$ yields

$$\tau_*^{-1}(\{\alpha\}) \simeq Sper(B \otimes_A k(\alpha)),$$

where $k(\alpha)$ denotes the real closed field attached to α and '$\simeq$' means 'homeomorphic'. In our present situation, $B \otimes_A k(\alpha)$ turns out to be a finite-dimensional $k(\alpha)$-algebra, hence $Sper(B \otimes_A k(\alpha))$ and the fiber over α are finite sets.

The final notion we need is that of the signature of a bilinear form φ on a finitely generated A-module M at a point $\alpha \in Sper\, A$: $sgn_\alpha(\varphi)$. By definition $sgn_\alpha\varphi := sgn(\varphi \otimes_A k(\alpha))$ where, of course, the scalar extension $\varphi \otimes_A k(\alpha)$ is defined on the finite-dimensional $k(\alpha)$-vectorspace $M \otimes_A k(\alpha)$ and the unadorned *sgn* is the usual signature of $k(\alpha)$, cf. [BCR, Ch. 15].

After all these preparations we can state

Theorem 1.1 *(The trace formula) Under the hypothesis above the following formula holds:*

$$sgn_\alpha(tr_{B|A})_*(\varphi) = \sum_{\substack{\beta|\alpha \\ \beta \in Sper\, B}} sgn_\beta(\varphi)$$

Proof. We start off from the situation $A \longrightarrow B$, M a B-module equipped with the B-form φ and the A-form $tr_*(\varphi) = (tr_{B|A})_*(\varphi)$. To calculate $sgn_\alpha(tr_{B|A})_*(\varphi))$ we have to carry out a scalar extension relative to the morphism $A \longrightarrow k(\alpha)$ and we then face the following situation: $k(\alpha) \longrightarrow B \otimes_A k(\alpha)$, $M \otimes_A k(\alpha)$ considered as a $B \otimes_A k(\alpha)$-module equipped with two bilinear forms:

1) the $B \otimes_A k(\alpha)$-valued form $\hat{\varphi}$ defined by

$$\hat{\varphi}(m \otimes x, n \otimes y) := \varphi(m,n) \otimes xy$$

and

2) the scalar extension $tr_*(\varphi) \otimes_A 1$ which is $k(\alpha)$-valued.

To facilitate the notations we set

$$\hat{M} := M \otimes_A k(\alpha), \hat{B} := B \otimes_A k(\alpha), \widehat{tr_*(\varphi)} := tr_*(\varphi) \otimes_A 1.$$

Being a finite-dimensional $k(\alpha)$-algebra, $\hat{B}$ decomposes uniquely into a direct sum of ideals B_i, which are local $k(\alpha)$-algebras ($\oplus$: $k(\alpha)$-direct sums):

$$\hat{B} = \bigoplus_{i=1}^{r} B_i, B_i = e_i B, e_i \text{ (indecomposable) idempotents.}$$

Since $\hat{M}$ is a $\hat{B}$-module it splits correspondingly:

$$\hat{M} = \bigoplus_{i=1}^{r} M_i, M_i = e_i \hat{M}$$

Since $M_i = \epsilon_i \hat{M}$, distinct M_i's are orthogonal relative to $\hat{\varphi}$. Set $\varphi_i = \hat{\varphi}_{|M_i}$. Then φ_i is B_i–valued, and we claim $\widehat{tr_*(\varphi)}_{|M_i} = (tr_{B_i|k(\alpha)})_*(\varphi_i)$. To prove this first note $\widehat{tr_*(\varphi)} = (tr_{\hat{B}|k(\alpha)})_*(\hat{\varphi})$, which is due to the compatibility of trace map with scalar extensions. Consequently, $\widehat{tr_*(\varphi)}_{|M_i} = (tr_{\hat{B}|k(\alpha)})_*(\varphi_i) = (tr_{B_i|k(\alpha)})_*(\varphi_i)$, since B_i is a direct ideal summand of $\hat{B}$.

From $\widehat{tr_*(\varphi)} = (tr_{\hat{B}|k(\alpha)})_*(\hat{\varphi})$ we further get that the M_i's also provide an orthogonal decomposition relative to $\widehat{tr_*(\varphi)}$. Putting the information together we have derived so far:

$$(*) \quad sgn_\alpha(tr_*(\varphi)) = sgn\ \widehat{tr_*(\varphi)} = \sum_{i=1}^{r} sgn\ (tr_{B_i|k(\alpha)})_*(\varphi_i)$$

Setting $C = B_i, N = M_i, \psi = \varphi_i, k = k(\alpha)$ we are dealing with a finite-dimensional local k–algebra C and a C–form ψ living on a finitely generated C–module N. We have to calculate $sgn\ (tr_{C|k})_*(\varphi)$.

We start by observing that in this case N is in fact a finitely generated projective C–module since we started out from a projective B–module and all the constructions above preserve projectivity. Since C is local, N is a free C–module. Let $\mathcal{M}$ denote the maximal ideal of C. Then the reduction $\overline{\varphi} = \varphi \otimes_C 1$ is a bilinear form on $\overline{M} = M \otimes_C C/_{\mathcal{M}} = M/_{\mathcal{M}M}$ which is a finite–dimensional vector space over the field $C/_{\mathcal{M}} = \overline{C}$. We want to prove $sgn\ (tr_{C|k})_*(\varphi) = sgn\ (tr_{\overline{C}|k})_*(\overline{\varphi})$. In our case, $char\ C/_{\mathcal{M}} \neq 2$, so $\overline{\varphi}$ has a decomposition $\overline{\varphi} = \hat{\varphi}_1 \perp \hat{\varphi}_2, \hat{\varphi}_1$ non–degenerate and $\hat{\varphi}_2$ a Null–form: $\hat{\varphi}_2 =< 0, \ldots, 0 >$. Using [Ba, (3.4), p. 11] we obtain a C–decomposition of $(N, \varphi) : \varphi = \varphi_1 \perp \varphi_2$ with φ_1 free and non singular and φ_2 having all its values in $\mathcal{M}$. Now, by [Ba, (3.5), p. 13], φ_1 admits an orthogonal basis, and therefore we are facing the following situation:

$$\varphi =< x_1 >\perp \ldots \perp< x_n >\perp \varphi_2, \varphi_2(N, N) \subset \mathcal{M}, x_i \in C^*.$$

Clearly $\overline{\varphi} =< \overline{x}_1 >\perp \ldots \perp< \overline{x}_n > \perp$ Null–form, and since the trace map $tr_{C|k}$ vanishes on $\mathcal{M} = rad\ C$ our last claim will follow in general once it is proved in the special case $\varphi =< x >, x \in C^*$.

We use the fact that C has a 'Wedderburn–decomposition' $C = C_0 \oplus \mathcal{M}, C_0$ a subalgebra (in fact a subfield), clearly satisfying $C_0 \simeq \overline{C}$. By using $tr_{C|k}(\mathcal{M}) = 0$, and $(tr_{C|k}) |_{C_0} = [C : C_0] \cdot tr_{C_0|k}$ one easily checks $sgn\ tr_{C|k} < x >= sgn\ tr_{\overline{C}|k} < \overline{x} >$ which then, as said above, implies our general claim.

Going back to the equation (*) we now have to compute $sgn\ (tr_{\overline{B}_i|k(\alpha)})_*(\overline{\varphi}_i)$. Since $k(\alpha)$ is real closed there are just two cases: $\overline{B}_i = k(\alpha)$ and $\overline{B}_i = k(\alpha)(\sqrt{-1})$. In the second case, $\overline{\varphi}_i$ is a torsion form, if considered in $W(\overline{B}_i)$, and so is $(tr_{\overline{B}_i|(k(\alpha))})_*(\varphi)$ in $W(k(\alpha)) \simeq \mathbf{Z}$. Hence $sgn\ (tr_{\overline{B}_i|k(\alpha)})_*(\overline{\varphi}_i) = 0$. In the first case, $(tr_{\overline{B}_i|k(\alpha)})_*(\overline{\varphi}_i) = \overline{\varphi}_i$, and using this we get the following refinement of $(*)$:

$$(**) \quad sgn_\alpha(\varphi) = \sum_{\overline{B}_i = k(\alpha)} sgn(\overline{\varphi}_i)$$

We next interpret the right hand side of $(**)$ by using *Sper* B. As proved in [C–R, prop. 4.3, p. 40], see also [Be], the natural morphism $B \longrightarrow B \otimes_A k(\alpha)$ induces a homeomorphism (= bijection in our case) between the fiber of *Sper* $B \longrightarrow$ *Sper* A over α and $Sper(B \otimes_A k(\alpha))$. From the above decomposition $\hat{B} = \oplus B_i$ one derives that the points β_i extending α corresponds in a 1–to–1 manner to those of the B_i satisfying $\overline{B}_i / \mathcal{M}_i = k(\alpha)$. More precisely, if such a B_i is given the associated β_i is defined as the homomorphism (pr = projection):

$$(* * *) \quad B \longrightarrow B \otimes_A k(\alpha) \xrightarrow{pr} B_i \longrightarrow B_i / \mathcal{M}_i = k(\alpha) = k(\beta_i)$$

Here and in the sequel we consider $\alpha, \beta, \ldots$ as maps. We next have to compute $(M \otimes_B k(\beta), \varphi \otimes_B k(\beta))$ if $\beta = \beta_i$ for some extension β_i. Since $k(\beta) = k(\alpha)$ we are studying the following situation:

$$\begin{tikzcd} M \arrow[d, dash] & \\ B \arrow[r, "\beta"] & k(\alpha) \\ A \arrow[u, "\tau"] \arrow[ur, "\alpha"'] \arrow[ur, phantom, "|||" near start] & \end{tikzcd}$$

Using the $k(\alpha)$–isomorphism

$$(M \otimes_A k(\alpha)) \otimes_{B \otimes_A k(\alpha)} k(\alpha) \simeq M \otimes_B k(\alpha) = M \otimes_B k(\beta)$$

$$(m \otimes x) \otimes y \longmapsto m \otimes xy$$

which is a special instance of [C–E, prop. 2.1, p. 165] and taking into account the way β is composed $(* * *)$ we finally obtain:

$$(M \otimes_B k(\beta_i), \varphi \otimes_B k(\beta_i)) = (\overline{M}_i, \overline{\varphi}_i).$$

This shows that $sgn_{\beta_i}(\varphi) = sgn(\varphi \otimes_B k(\beta_i)) = sgn\ \overline{\varphi}_i$, and the proof is complete. □

In the literature, other linear forms $s : B \longrightarrow A$ are considered as well. In deriving a similiar transfer formula for such more general linear forms one has to cope with the following problem. Dealing with $sgn_\alpha s_*(\varphi)$ one decomposes $M \otimes_A k(\alpha)$, $B \otimes_A k(\alpha)$, $\varphi \otimes_A 1$, $s \otimes_A 1 : \hat{B} \longrightarrow k(\alpha)$ into pieces $M_i, B_i, \varphi_i, s_i : B_i \longrightarrow k(\alpha)$ and studies $s_{i*}(\varphi_i)$. But, in the case of $sgn_{\beta_i}(\varphi), \beta_i$ belonging to B_i, one is concerned with $\overline{M}_i$ and $\overline{\varphi}_i$. Hence, $s_{i*}(\varphi_i)$ has to be compared to $\overline{\varphi}_i$. As shown above, $\varphi_i =< x_1 >\perp \ldots \perp< x_n >\perp \psi, x_i \in C^*, \psi(N, N) \subset \mathcal{M}_i$. From ${B_i}\big/{\mathcal{M}_i} = k(\alpha), \frac{1}{2} \in B_i$ and the fact that B_i is a finite dimensional local $k(\alpha)$–algebra we get $x_i = a_i \cdot \epsilon_i^2, a_i \in k(\alpha)^*, \epsilon_i \in C^*$. Thus $\varphi_i =< a_1 >\perp \ldots \perp< a_n >\perp \psi$. Invoking Frobenius–reciprocity as in [K2, (1.1), p. 169], which is valid for arbitrary linear forms, we obtain

$$s_{i*}(\varphi_i) = (s_{i*} < 1 >) \cdot < a_1, \ldots, a_n > \perp s_{i*}(\psi)$$

and $\overline{\varphi}_i =< a_1, \ldots, a_n >\perp$ Null–form. Thus

$$sgn\ s_{i*}(\varphi_i) = (sgn\ s_{i*} < 1 >) \cdot sgn\ \overline{\varphi}_i + sgn\ s_{i*}(\psi).$$

Setting $m(\beta_i, \alpha) = sgn\ s_{i*} < 1 > \in \mathbf{Z}$ we finally get

$$sgn_\alpha s_*(\varphi) = \sum_{\beta|\alpha} m(\beta, \alpha) sgn_\beta(\varphi) + \sum_{\beta|\alpha} sgn\ s_{\beta *}(\psi_\beta).$$

In order to derive a transfer formula in the sense of Knebusch, cf. [K2, K3], one needs assumptions implying the vanishing of the second sum on the right hand side. If we either assume that $B \mid A$ is étale, forcing $B \otimes_A k(\alpha)$ to be separable, or that φ is non–degenerate then ψ does not even occur. If we assume that $s \otimes_A 1$ vanishes on the radical of $B \otimes_A k(\alpha)$ for all $\alpha \in Sper\ A$ then $s_{i*}(\psi)$ is a Null–form; e.g. this happens if s is of the type $s(x) = tr_{B|A}(xa)$ for some fixed $a \in B$. Summarizing we get

Proposition 1.2 *In each of the following cases*

- *$B \mid A$ étale,*
- *only non–degenerate forms are considered,*
- *$s \otimes_A 1 = 0$ on $Nil(B \otimes_k k(\alpha))$ for each $\alpha \in Sper\ A$*

there are integers $m(\beta, \alpha)$, for every $\alpha \in Sper\ A, \beta \in Sper\ B$ such that $\beta \mid \alpha$, allowing the following transfer formula:

$$sgn_\alpha s_*(\varphi) = \sum_{\beta|\alpha} m(\beta,\alpha) sgn_\beta(\varphi)$$

for any one of forms φ being considered.

The case of an étale extension is dealt with in [M, (4.11)].

Note that $m(\beta,\alpha) = 1$ if $s = tr_{B|A}$.

If we are neither in the first or third case of the proposition the restriction to non-degenerate forms is essential. To see this consider Knebusch's example of a Frobenius algebra in [K3, p. 177]:

$$A = \mathbb{R}, B = \mathbb{R}[t], t^2 = 0, \{1,t\} \ \mathbb{R}\text{-basis of } B.$$

The unique $\alpha \in Sper\ A$ has a unique extension β to B. Consider the A-linear form s given by $s(1) = 0, s(t) = 1$. Then $s_* < t > \simeq < 1, 0 >$, hence a formula $sgn_\alpha s_*(\varphi) = m \cdot sgn_\beta(\varphi)$ cannot exist.

As an example where the last condition of prop. 1.2 is deliberately violated is the famous Eisenbud-Levine formula, cf. [E-L]. In that case one studies linear forms $s : B \to k$, k a field with $s \neq 0$ on the Nilradical of B, and such forms are needed for the application in mind.

If $(B \mid A)$ is a Frobenius extension then Knebusch's transfer formula for signatures, cf. [K3],

$$\sigma(s_*\varphi) = \sum_{\tau|\sigma} n(\tau)\tau(\varphi)$$

can be derived from our formula above as he has kindly pointed out to us. To this end, let $Sign\ A$ denote the set of signatures $\sigma : W(A) \to \mathbf{Z}$ as in [K5, §5]. It is known that the natural map $Sper\ A \to Sign\ A$, $\alpha \mapsto \sigma_\alpha$ is surjective ($\alpha \in Sper\ A$ induces $A \to k(\alpha)$, hence $\sigma_\alpha : W(A) \to W(k(\alpha)) = \mathbf{Z}$), cf. [K5, §5] or [K4, §§1,3]. In [K3, (1.1), p. 169] it was shown that there is a unique choice of the multiplicities $n(\tau)$. They are all strictly positive in the case of $s = tr_{B|A}$, cf. [K2, (3.4), p. 72].

To derive the trace formula for signatures let $\sigma = \sigma_\alpha$ and fix α. Then, for any non-degenerate form: $\sigma(\varphi) = sgn(\varphi \otimes_A k(\alpha)) = sgn_\alpha(\varphi)$. If $\beta \mid \alpha$ then $\sigma_\beta \mid \sigma_\alpha$. If $\tau \mid \sigma$ then $n(\tau) = \sum_\beta m(\beta,\alpha)$, summing up all β's with $\sigma_\beta = \tau$. With this definition we obviously get from our trace formula

$$\sigma(s_*(\varphi)) = \sum_{\tau|\sigma} n(\tau)\tau(\varphi).$$

If $s = tr_{B|A}$ then $m(\beta, \alpha) = 1$ hence $n(\tau) \geq 0$. Moreover, $n(\tau) > 0$ in the unique trace formula, as cited above. Therefore, if $\sigma = \sigma_\alpha$ then any extension $\tau \mid \sigma$ arises from an extension $\beta \mid \alpha$, i.e. $\tau = \sigma_\beta$.

2 Counting real points

In this section we are going to apply the trace formula to the counting of real points on 0-dimensional affine varieties over real closed fields. We will use the book [BCR] as the basic reference for real algebraic geometry.

Let R denote a real closed field. It is one of the basic computational tasks in real algebraic geometry to decide whether a given semi algebraic $S \subset R^n$ is empty or not. Here, wlog, S may be thought to be given as the set of solutions in R^n of a system of the following type:

$$F_1(X_1,\ldots,X_n) = 0, \ldots F_r(X_1,\ldots,X_n) = 0,$$

$$G_1(X_1,\ldots,X_n) > 0, \ldots G_s(X_1,\ldots,X_n) > 0$$

with polynomials $F_1,\ldots,G_s \in R[X_1,\ldots,X_n]$. The decision about '$S = \emptyset$' or '$S \neq \emptyset$' should be made by using the coefficients of the polynomials involved in the above presentation of S.

Our problem can be rephrased as follows. The equations $F_1 = 0,\ldots,F_r = 0$ define an affine variety V with coordinate ring $R[X_1,\ldots,X_n] \big/ (F_1,\ldots,F_r)$, and the polynomials $G_1,\ldots,G_s$ gives rise to regular functions $g_1,\ldots,g_s$ on V (g_i being the coset of G_i in $R[V]$). In case $K \supset R$ is a field extension, $V(K)$ denote the set of points of V with coordinates in K.

In this setting S can be described as follows:
$S = \{x \in V(R) \mid g_1(x) > 0,\ldots,g_s(x) > 0\}$.

If V is a 0-dimensional variety, i.e. Krull-dim $R[V] = 0$ or, equivalently, $\#V(R(\sqrt{-1})) < \infty$, we will attach to this representation of S a quadratic form φ over R such that

$$\#S = \frac{1}{2^s} sgn\ \varphi.$$

To count real points by calculating the signature of appropriate quadratic forms is a topic really begun in the last century by Borchardt, Jacobi, Sylvester and, first of all, Hermite. They studied the case $V = \mathbb{R}$ and $V = \mathbb{R}^2$, and their method is often referred to as the Hermite-Sylvester method. A very comprehensive account of this approach, complete up to about 1939, can be found in the reprint of a survey by Krein and Naimark, cf. [Kr-N]. Also the books of Knebusch-Scheiderer [K-S] and Benedetti-Risler [B-R, p. 17 ff] are recommended for further information.

Our general treatment of arbitrary 0-dimensional varieties stems from a geometric interpretation of the trace formula of §1. Quite astonishingly, P. Peddersen in a nearly simultaneous and completely independent study also introduced the same quadratic forms to achieve the joint goal: counting real zeros on zero–dimensional varieties, cf. [P1]. A brief account of our method has already been published in [Be2].

Now, let V be a 0–dimensional affine variety over the real closed field R with coordinate ring $R[V]$. Each real point $x \in V(R)$ gives rise to the evaluation map $e_x : R[V] \to R$, $f \mapsto f(x)$, which in turn yields the point $\alpha_x \in Sper\ R[V]$ defined as $\alpha_x = (\mathcal{M}_x, R_+)$, $\mathcal{M}_x = ker\ e_x$. From $dim\ R[V] = 0$ one concludes $Sper\ R[V] = \{\alpha_x \mid x \in V(R)\}$. If $\varphi =< f_1, \dots, f_s >$ is any diagonizable quadratic form over $R[V]$ (as $\frac{1}{2} \in R[V]$ no difference is made between bilinear and quadratic forms) then $\varphi_x :=< f_1(x), \dots, f_s(x) >$ is a form over R and we have

$$sgn_{\alpha_x}\varphi = sgn\ < f_1(x), \dots, f_s(x) > .$$

In our geometric context, the trace formula of §1 now reads

Proposition 2.1 *For any* $f_1, \dots, f_s \in R[V]$

$$sgn\ ((tr_{R[V]|R})_* < f_1, \dots, f_s >) = \sum_{x \in V(R)} sgn\ < f_1(x), \dots, f_s(x) > .$$

We are next going to specialize this geometric trace formula. If $s = 1$, $f_1 = 1$, then $sgn\ tr_* < 1 >= \#V(R)$. For given $f_1, \dots, f_s \in R[V]$ we form the 'scaled Pfisterform' $\varphi =< \prod_1^s f_i > \cdot \ll f_1, \dots, f_s \gg$. An immediate calculation shows

$$sgn\ \varphi_x = \begin{cases} 2^s, & \text{if } f_1(x) > 0, \dots, f_s(x) > 0 \\ 0, & \text{otherwise} \end{cases}$$

Summarizing these two cases we get

Proposition 2.2 *a)* $sgn((tr_{R[V]|R})_* < 1 >) = \#V(R)$,

b) $sgn((tr_{R[V]|R})_* < \prod_1^s f_i >\ll f_1, \dots, f_s \gg) = 2^s \cdot \#\{x \in V(R) \mid f_1(x) > 0, \dots, f_s(x) > 0\}$

Clearly, the right–hand sides in the last proposition give the number of points one is interested in. However, if it comes to actual computation one necessarily has to deal with the left–hand sides of the above equations in an explicit and efficient way. In the following we will outline one possible way to cope with the trace forms themselves. There are clearly other methods and best efficiency is not claimed. In forthcoming papers by Pedersen, Roy and Spzirglas [P3] the complexity of all the algorithms involved will be discussed in great detail.

Let V be given by a set of polynomial equations

$$F_1(X_1, \dots, X_n) = 0. \dots, F_r(X_1, \dots, X_n) = 0,$$

i.e. by setting $\mathfrak{a} = (F_1, \dots, F_r) \lhd R[X_1, \dots, X_n]$ we have $R[V] = {R[X_1, \dots, X_n]}\big/_{\mathfrak{a}}$. Note that $\mathfrak{a}$ is not assumed to be a radical ideal.

The elements f_i are represented by polynomials G_i, $i = 1, \dots, s$.

Let k be the field obtained by adjoining to $\mathbb{Q}$ all the coefficients of $F_1, \dots, F_r$ and, in the second case of prop. 2.2, of $G_1, \dots, G_s$. We set $\mathfrak{a}_0 := (F_1, \dots, F_r) \lhd k[X_1, \dots, X_n]$. A careful reading of the following reasoning will show that all the necessary arithmetic operations can be carried out in k. Note $\mathfrak{a} = \mathfrak{a}_0 R[X_1, \dots, X_n]$.

In prop. 2.2 only forms over $R[V]$ admitting an orthogonal basis occur. So, it is enough to deal with the case $\psi = (tr_{R[V]|R})_* < h >$, where $h = G + \mathfrak{a}$, $G \in k[X_1, \dots, X_n]$. To derive a matrix for ψ one needs an R-basis of $R[V]$. This can be achieved by obtaining a Gröbner–basis for $\mathfrak{a}_0$ using the Buchberger algorithm which is performed inside k. From such a Gröbner–basis of $\mathfrak{a}_0$ one gets a k-basis of ${k[X_1, \dots, X_n]}\big/_{\mathfrak{a}_0}$ which remains a R-basis of ${R[X_1, \dots, X_n]}\big/_{\mathfrak{a}}$ under scalar extension. To see the details of this argumentation one may consult e.g. [Bu]. We also use this reference as a source for all what is needed about Gröbner Bases and the Buchberger algorithm.

We will proceed differently to get finally an R-basis on an appropriate R-algebra which has more structure and allows an easier way of determining the signature. However, costs are due for deriving this other algebra.

Set $A = R[V]$. Being a finite-dimensional R-algebra, A admits a Wedderburn decomposition

$$A = J \oplus \overline{A}, J = \text{ Nilradical, } \overline{A} \text{ subalgebra}$$

where the natural projection $\pi : A \to A/_J = A_{red}$ induces an R-isomorphism $\overline{A} \simeq A_{red}$. We now resume arguments presented in the proof of the trace formula. Setting $h = u \oplus \overline{h}$, $u \in J$, $\overline{h} \in \overline{A}$ we get

$$sgn\ (tr_{A|R})_* < h >= sgn\ (tr_{\overline{A}|R})_* < \overline{h} > .$$

Being only interested in signatures we should therefore continue with the determination of $sgn\ (tr_{\overline{A}|R})_* < \overline{h} >$. To this end we have to calculate $\sqrt{\mathfrak{a}}$ out of $\mathfrak{a}$ since $\overline{A} \simeq {R[X_1, \dots, X_n]}\big/_{\sqrt{\mathfrak{a}}}$. To determine $\sqrt{\mathfrak{a}}$ we use the following observation of Seidenberg, cf. [Se]. First note that for any $i = 1, \dots, n$

$$\mathfrak{a}_0 \cap k[X_i] \neq (0),$$

namely, if not so, then, because of $\mathfrak{a} \cap k[X_1, \dots, X_n] = \mathfrak{a}_0$, we had an injection (over k) $k[X_i] \to {R[X]}\big/_{\mathfrak{a}}$ which is impossible since ${R[X]}\big/_{\mathfrak{a}}$ is an algebraic k-algebra. Now let $g_i \neq 0$ be arbitrarily chosen in $\mathfrak{a}_0 \cap k[X_i], i = 1, \dots, n$. Using gcd-calculation inside

$k[X_i]$ we make g_i squarefree: $\tilde{g}_i = \frac{g_i}{gcd(g_i,g_i')}$. Then clearly $\tilde{g}_i \in \sqrt{\mathfrak{a}}, i = 1,\dots,n$, and, as Seidenberg noticed,

$$\sqrt{\mathfrak{a}} = (\mathfrak{a}, \tilde{g}_1, \dots, \tilde{g}_n).$$

To prove this statement first note that each R-algebra $R[X_i] \big/ _{(\tilde{g}_i)}$ is separable as *char* $R = 0$. Consequently, also $B = \bigotimes_{i=1}^{n} R[X_i] \big/ _{(\tilde{g}_i)}$ is separable, i.e. a finite product of field extensions of R. Obviously there is a natural epimorphism

$$B \twoheadrightarrow R[X_1,\dots,X_n] \big/ _{(\tilde{g}_1,\dots,\tilde{g}_n)} \twoheadrightarrow R[X_1,\dots,X_n] \big/ _{(\mathfrak{a},\tilde{g}_1,\dots,\tilde{g}_n)}$$

showing the latter algebra to be separable, hence $(\mathfrak{a}, \tilde{g}_1, \dots, \tilde{g}_n)$ is a radical ideal. This proves the claim. The same argument applied to $\mathfrak{a}_0$ shows $\sqrt{\mathfrak{a}_0} = (\mathfrak{a}_0, \tilde{g}_1, \dots, \tilde{g}_n)$.

In order to calculate $tr_{\overline{A}|R} < \overline{h} >$ one needs an R-basis of $\overline{A}$. So far, there seems to be no advantage to use $\overline{A}$ instead of A. However, it is the so called (folklore) Shape Lemma, cf. [GiTrZ], that guarantees a distinguished set of generators of $\sqrt{\mathfrak{a}_0}$, hence of $\sqrt{\mathfrak{a}} = \sqrt{\mathfrak{a}_0} \cdot R[X_1,\dots,X_n]$. Thus we get a nice basis for $\overline{A}$.

Shape Lemma. Let k be a infinite perfect field, $\mathfrak{b}$ any 0–dimensional radical ideal in $k[X_1,\dots,X_n]$. Then, possibly only after a linear change of coordinates, i. e. $Y_i = X_i, i = 1,\dots,n-1, Y_n = X_n + \sum_1^{n-1} X_i \cdot t^i$, for some $t \in k$, there are polynomials $g_1,\dots,g_{n-1}, g \in k[T], g \neq 0$, square free, degree $g_i <$ degree g such that

$$\mathfrak{b} = (Y_1 - g_1(Y_n), \dots, Y_{n-1} - g_{n-1}(Y_n), g(Y_n))$$

Proof: Set $V = \{x \in \overline{k}^n \mid F(x) = 0 \text{ for all } F \in \mathfrak{b}\}$ where $\overline{k}$ = algebraic closure. Since *dim* $\mathfrak{b} = 0$, V is finite. We consider the projection $pr : V \to \overline{k}, (x_1,\dots,x_n) \mapsto x_n$. Assume it to be injective, i.e., by definition, V to be in 'general position'. The Galois group $G = Gal(\overline{k} \mid k)$ operates naturally on V, and $pr(V)$ is the union of full conjugacy classes. This implies the existence of a square free polynomial (note: k perfect) $g \in k[T]$ such that $pr(V) = \{x \in k \mid g(x) = 0\}$. For each $x_n \in pr(V)$ let $(x_n^1, \dots, x_n^{n-1}, x_n)$ be the unique point in V over x_n. By using the Lagrange interpolation formula and making use of the G-action on V one actually gets polynomials $g_1,\dots,g_{n-1} \in k[X]$, degree $g_i <$ degree g such that, for all $x_n \in pr(V), (g_1(x_n),\dots,g_{n-1}(x_n), x_n)$ is the point over x_n. This shows that $\mathfrak{b}$ and $\mathfrak{b}' = (X_1 - g_1(X_n), \dots, X_{n-1} - g_{n-1}(X_n), g(X_n))$ have the same points in $\overline{k}$. Since $k[X_1,\dots,X_n] \big/ _{\mathfrak{b}'} \simeq k[X_n] \big/ _{(g(X_n))}$ is separable we see that also $\mathfrak{b}'$ is a radical ideal. Hence, $\mathfrak{b} = \mathfrak{b}'$ by Hilbert's Nullstellensatz.

In the case that $\mathfrak{a}$ is not in general position we have to adjust it. Using the above mentioned coordinate transformation one sees that all but finitely many t's will put $\mathfrak{a}$ into the desired general position. □

In our situation, if we had $\sigma\!\tau_0$ in general position, then $\overline{A} = R[X_1, \ldots, X_n \big/ (X_1 - g_1(X_n), \ldots, X_{n-1} - g_{n-1}(X_n), g(X_n)) \simeq R[T] \big/ (g(T))$ with $g \in k[T]$. Also, and this is crucial, $\overline{A}$ would admit the simple basis $1, T, \ldots, T^{N-1}$, $N =$ degree g. Relative to this basis the matrix for $(tr_{\overline{A}|R})_* < \overline{h} >$ shows additional features which allows a more accessible determination of its signature. Before turning to this point we must find the generators of $\sigma\!\tau_0$ as given in the Shape Lemma. It is the key point for our calculation that this set of generators allows a conceptual characterization as was first pointed out by Gianni and Mora, cf. [Gi–Mo]. Using the definition of a reduced (= minimal) Gröbner basis one readily verifies that, under the above conditions on $g_1, \ldots, g_{n-1}, g$,

$$X_1 - g_1(X_n), \ldots, X_{n-1} - g_{n-1}(X_n), g(X_n)$$

form the reduced Gröbner basis relative to the lexicographical order satisfying $X_1 > X_2 > \ldots > X_n$.

By the remarks above we derive the actual computation of $\sqrt{\sigma\!\tau_0}$ hence of $\overline{A} = R[X_1, \ldots, X_n] \big/ \sqrt{\sigma\!\tau}$. One first determines $g_i \in \sigma\!\tau_0 \cap k[X_i], g_i \neq 0, i = 1, \ldots, n$, e.g. by Gröbner bases techniques, cf. [Bu]. One then determines the reduced Gröbner basis of $(\sigma\!\tau_0, \tilde{g}_1, \ldots, \tilde{g}_n) = \sqrt{\sigma\!\tau_0}$ relative to the term ordering above. If this basis is not as expected then a random choice of the parameter t in the coordinate transformation will help to get this position and a second Gröbner basis computation will do the job.

Having done all this we are constructively given polynomials $g_1, \ldots, g_{n-1}, g \in k[X_n]$ and an R–isomorphism

$$\overline{A} \longrightarrow R[T] \big/ (g(T)) = \hat{A}. X_i \mapsto g_i(T), X_n \rightarrow T (i = 1, \ldots, n-1).$$

Next we use the standard basis $1, t, \ldots, t^{N-1}$ where $t = T + (g(T))$, $N =$ degree of g. Given $h(T) \in k[T]$ we have to determine the matrix of $(tr_{\hat{A}|R})_* < h(t) >$ relative to this basis. At the place (i, j) of this matrix we find

$$tr_{\hat{A}|R}(h(t)t^{i+j-2}) = \sum_{\substack{\alpha \in R(\sqrt{-1}) \\ g(\alpha)=0}} h(\alpha)\alpha^{i+j-2}.$$

Invoking the symmetric function theorem we get that the right hand side is a $\mathbf{Z}$–polynomial in the coefficients of $h(T)$ and $g(T)$. The chapter 4 of [P2] is devoted to a study of algorithms for evaluating symmetric functions. One may use those methods, but there is another way to determine $tr_{\hat{A}|R}(h(t)t^l)$. One expands the rational function $T\frac{h(T)g'(T)}{g(T)}$ in the formal power series field $k((T^{-1}))$ and passes to $\overline{k}((T^{-1})) \supset k((T^{-1}))$. Writing $g(T) = \prod(T - a)$ we get $h(T)\frac{g'(T)}{g(T)} = \sum_a \frac{h(T)}{T-a} = \sum_a \frac{h(\alpha)}{T-a} + H(T)$ for some polynomial $H \in \overline{k}[T]$. After multiplying by T we finally get: $T\frac{h(T)g'(T)}{g(T)} = TH(T) + \sum_a \frac{h(\alpha)}{1-aT^{-1}} = TH(T) + \sum_{l=0}^{\infty}(\sum_a h(a)a^l)T^{-l}$. Hence $tr_{\tilde{A}|R}(h(t)t^l)$ is the coefficient of T^{-l} in this expansion.

The resulting matrix is a so called Hankel matrix $H = (a_{i+j-2})_{i,j=1,\dots,N}$ built up from a sequence $a_0, \dots, a_{2N-2} \in k$. There are efficient methods for determing the signature of a Hankel matrix, e.g. by a theorem of Frobenius. To see the details one may consult [G] or [I].

So far we have outlined a method to compute $sgn\ (tr_{A|R})_* < h >$. This clearly implies the determination of $\#V(R) = sgn\ (tr_{A|R})_* < 1 >$. However in the case of the second statement of prop. 2.2 we would be forced to cope with as many as 2^s calculations of the type $sgn\ (tr_{A|R})_* < h_i >$. Even for fairly small values of s this would be beyond any feasible limit. In the next section we show that in fact the simultaneous inequalities $f_1 > 0, \dots, f_s > 0$ can be replaced by just a single one $h > 0$, and a single one can be handled as above. The reduction of the s inequalities to just one is not without expenses. So one should look for other methods. The fundamental B-K-R algorithm of [BKR] applies to several inequalities by an ingenious procedure using only one inequality in each step. For further reading one may turn to Pedersen's paper or the forthcoming ones by Pedersen, Roy and Spzirglas. Also, after the reduction to the univariate case $\tilde{A} = {}^{R[T]}\big/_{(g(T))}$ other methods are available as well, c.f. [GLRR].

3 The 0–dimensional case of the Bröcker–Scheiderer theorem

It was proved by C. Scheiderer [S] and L. Bröcker (unpublished) that in any n–dimensional affine variety V over a real closed field R a basic open set $S = \{x \in V(R) \mid f_1(x) > 0, \dots, f_r(x) > 0\}$ can in fact be described by at most $\overline{n}$ inequalities ($\overline{n} = max(1, n)$):

$$S = \{x \in V(R) \mid h_1(x) > 0, \dots, h_{\overline{n}}(x) > 0\}, h_1, \dots, h_{\overline{n}} \in R[V].$$

This amazing theorem is a real challenge to computation since all known proofs do not offer constructive methods to find $h_1, \dots, h_{\overline{n}}$.

In the sequel we consider the case of dim $V = 0$ and will propose an algorithm to find h_1 starting with a description of V and polynomials representing $f_1, \dots, f_r$.

So let V be described by polynomials as

$$V : F_1 = 0, \dots, F_s = 0$$

where $F_1(X_1, \dots, X_n), \dots, F_s(X_1, \dots, X_n) \in k[X_1, \dots, X_n]$ for some $k \subset R$. We also assume $f_1, \dots, f_r \in k[X_1, \dots, X_n]$. Finally we set $\mathfrak{a}_0 = (F_1, \dots, F_s) \lhd k[X_1, \dots, X_n], \mathfrak{a} = \mathfrak{a}_0 R[X_1, \dots, X_n]$. As explained in the last section, e.g. by Gröbner–bases techniques, we find the following generators of $\sqrt{\mathfrak{a}_0}$ (possibly after a coordinate transformation)

$$X_1 - g_1(X_n), \dots, X_{n-1} - g_{n-1}(X_n), g(X_n)$$

with $g_1, \dots, g \in k[X_n]$, g square–free, $deg\ g_i < deg\ g, (i = 1, \dots, n-1)$. Setting $N(g) = \{\alpha \in R \mid g(\alpha) = 0\}$ we find an isomorphism

$$\Phi : V(R) \to N(g), (x_1, \dots, x_n) \mapsto x_n$$

sending our regular functions $f_1(X_1, \dots, X_n), \dots, f_r(X_1, \dots, X_n)$ into the univariate polynomials $f_i(g_1(T), \dots, g_{n-1}(T), T) = \overline{f}_i \in k[T], i = 1, \dots, r$. After this transformation we are concerned with the set

$$(3.1) \quad \overline{S} = \{\alpha \in R \mid g(\alpha) = 0, \overline{f}_1(\alpha) > 0, \dots, \overline{f}_r(\alpha) > 0\}.$$

Suppose we have found $h(T)$ such that $\overline{S} = \{\alpha \in R \mid g(\alpha) = 0, h(\alpha) > 0\}$. Then $\{x \in V(R) \mid f_1(x) > 0, \dots, f_r(x) > 0\} = \{x \in V(R) \mid h(x_n) > 0\}$. Hence, we have to settle the case (3.1) under the assumption that g is squarefee. We first want to point out that, in our case, the Bröcker-Scheiderer result is an immediate consequence of the geometric trace formula and some simple quadratic form theory. Setting $\varphi = <\Pi\overline{f_i}> \cdot \ll \overline{f_1}, \dots, \overline{f_r} \gg$ we will find $h \in R[T]$ with $\varphi = 2^{r-1} \cdot <h> \ll h \gg$ over

$A = R[T] \big/ (g(T))$. As g is squarefree we have $A = \prod R \times \prod R(\sqrt{-1})$ where each factor R corresponds to a point $x \in N(g)$. If some $\overline{f_i}(x) = 0$ then $\varphi_x =$ Null-form of dimension 2^r and if all $\overline{f_i}(x) \neq 0$ either φ_x hyperbolic or $\varphi_x \simeq 2^r \times <1>$. Since quadratic form theory over R and $R(\sqrt{-1})$ is completely known one easily finds $h \in R[T]$ as desired.

The algorithms we are going to propose will work in the following two cases:

a) $R = \mathbb{R}$,

b) $R = \mathbb{R}\{\epsilon_1\}\ldots\{\epsilon_n\}$, an iterated Puiseux-series field with ϵ_i infinitesimal small relative to $\mathbb{R}\{\epsilon_1\}\ldots\{\epsilon_{i-1}\}$.

For definitions cf. [BCR, (1.2.3), p. 10].

1^{st} case: $R = \mathbb{R}$

Once the binary case $r = 2$ in $(*)$ is settled a recursive procedure will cover the general case. So, consider the situation

$$g(\alpha) = 0, f_1(\alpha) > 0, f_2(\alpha) > 0, g, f_1, f_2 \in k[T].$$

and notice that we do not know the zeros of g in k. The following reasoning will make use of the $\alpha \in N(g)$ only on a conceptual level; the algorithm itself entirely deals with g, f_1, f_2, i.e. their coefficients. The set $N(g)$ being finite implies that there are polynomials, say,

$$sgn\ f_1, sgn\ f_2 \in R[T]$$

such that $(sgn\ f_i)(\alpha) = sgn(f_i(\alpha)), i = 1,2$ holds. Now, set $H = sgn\ f_1 \cdot sgn\ f_2$ $(1 + sgn\ f_1 + sgn\ f_2)$ then one readily verifies for all $\alpha \in N(g)$:

$$(3.2) \quad H(\alpha) > 0 \Longleftrightarrow f_1(\alpha) > 0, f_2(\alpha) > 0.$$

Thus, it remains to compute $sgn\ f_i$ as defined above. As a matter of fact, we don't know of any way to do that. If one could compute the square root of positive functions in $A = R[T] \big/ (g(T))$ then an application to $\sqrt{f^2}$ would help to find $sgn(f) \in A$. But, no method is known to us. However, looking at (3.2) and the definition of H one observes that any pair of sufficiently close approximations of $sgn(f_i), i = 1,2$ will serve as well. In fact, if for $i = 1, 2, s_\epsilon(f_i)$ denote a polynomial with

$$| s_\epsilon(f_i) - sgn\ f_i | < \epsilon \quad \text{on} \quad N(g) \quad \text{and} \quad s_\epsilon(f_i)(\alpha) = 0 \quad \text{whenever} \quad f_i(\alpha) = 0$$

and if $0 < \epsilon < \frac{1}{3}$ then setting $H_\epsilon = s_\epsilon(f_1) \cdot s_\epsilon(f_2) \cdot (1 + s_\epsilon(f_1) + s_\epsilon(f_2))$ we also get on $N(g)$:

$$(3.3) \quad H_\epsilon > 0 \Longleftrightarrow f_1 > 0, f_2 > 0$$

Consequently, given $f \in \mathbb{R}[T]$, we have to find an approximation of $sgn(f)$ on $N(g)$ by a polynomial $s_\epsilon(f) \in \mathbb{R}[T]$. This will be basically achieved by a global Newton method applied to the ring $A_0 = \mathbb{R}[T] \big/ (g(T))$, the latter being considered as a ring of functions on $N(g)$. To display the basic idea we assume that all of the following operations can be carried out in A_0. We are going to write down the Newton sequence for the equation $X^2 - 1 = 0$ in A_0 starting with the initial value $f_0 = f$, i.e. we get (f_k) where

$$f_{k+1} = \frac{1}{2}(f_k + \frac{1}{f_k}), \quad f_k \in A_0.$$

Looking at the parabola defined by $y = x^2 - 1$ and taking the geometric interpretation of the Newton method into account one readily checks that (f_k) converges to $sgn\ f$ on $N(g)$.

However, to carry out this idea we have to cope with the situation that f or some f_k are not units in A_0, i.e. that some f_k are not relatively prime. That this can happen accounts for a more careful approach. The basic idea will be kept nevertheless.

Proposition 3.1 *Given $f, g \in \mathbb{R}[T]$ one can construct a sequence of polynomials (T_k) in $\mathbb{R}[T]$, without knowing the zero-set $N(g)$, such that*

(i) $\lim_{k \to \infty} F_k(\alpha) = sgn\ f(\alpha)$ *for every* $\alpha \in N(g)$,

(ii) $F_k(\alpha) = 0$ *for every* $\alpha \in N(g)$ *satisfying* $f(\alpha) = 0$.

Proof: <u>1st case:</u> f, g relatively prime. We are going to construct two sequences $(g_k), (F_k), k \geq 0$, in $\mathbb{R}[T]$ subject to

a) $g_0 = g, g_k \mid g_{k-1}, N(g_k) = N(g)$,

b) $F_0 = f, F_k$ and g_k relatively prime,

$$F_{k+1}(\alpha) = \frac{1}{2}\left(F_k(\alpha) + \frac{1}{F_k(\alpha)}\right) \quad \text{for all} \quad \alpha \in N(g).$$

Since, by assumption, there is no $\alpha \in N(g)$ with $f(\alpha) = 0$, condition (ii) is empty, and so the sequence (F_k) has the desired properties. Assume g_k, F_k are constructed. Then $f_k := F_k + (g_k) \in \mathbb{R}[T] \big/ (g_k) = A_k$ is a unit in A_k. Hence we can form $\tilde{f}_{k+1} = \frac{1}{2}(f_k + \frac{1}{f_k})$ in A_k. Note that f_k^{-1} can be computed by the Euclidean algorithm applied to F_k and g_k in $\mathbb{R}[T]$. Choose any $F_{k+1} \in \mathbb{R}[T]$ representing $\tilde{f}_{k+1}$ in A_k. Set $g_{k+1} = g_k / gcd(F_{k+1}, g_k)$. Since g_k, as a divisor of g, is squarefree we get that F_{k+1} and g_{k+1} are relatively prime.

Clearly, $N(g_{k+1}) \subseteq N(g_k)$, but any $x \in N(g_k)\backslash N(g_{k+1})$ would satisfy $g_k(\alpha) = 0 = F_{k+1}(\alpha)$ implying $f_k^2(\alpha) + 1 = 0$ which is impossible in view of $\alpha \in \mathbb{R}$, $f_k \in \mathbb{R}[T]$. Hence, the 1^{st} case is settled.

<u>2^{nd} **case:**</u> Since g is squarefree the polynomials $\tilde{g} = g/gcd(g,f)$ and f are relatively prime. So we find a sequence $(\tilde{F}_k)$ doing the job on $N(\tilde{g})$ which is $N(g)\backslash N(f)$. Using again that $gcd(\tilde{g},f) = 1$ we find by the Euclidean algorithm a polynomial, say, $(f,g)^{-1} \in k[T]$ satisfying $(f,g)^{-1} \cdot f \equiv 1 \quad mod \quad \tilde{g}$. Now set

$$\chi = (f,g)^{-1} \cdot f$$

then on $N(g)$:

$$\chi(\alpha) = \begin{cases} 1 & f(\alpha) \neq 0, \quad i.e. \quad \alpha \in N(\tilde{g}) \\ 0 & f(\alpha) = 0, \quad i.e. \quad \alpha \notin N(\tilde{g}), \end{cases}$$

so χ is the characteristic function of $N(\tilde{g})$. Hence, setting $F_k := \chi \cdot \tilde{F}_k$, we get the desired sequence. □

It remains to decide when to stop the sequence (F_k) in order to get the approximation

$$| F_k(\alpha) - sgn\ f(\alpha) | < \frac{1}{3} \quad \text{for} \quad \alpha \in N(g).$$

On $N(g) \cap N(f)$ we have $F_k(\alpha) = 0$, so no problem arises. On $N(g)\backslash N(f)$ we have $F_{k+1}(\alpha) = \frac{1}{2}(F_k(\alpha) + \frac{1}{F_k(\alpha)})$. In particular, $| F_k(\alpha) | \geq 1$ for $k \geq 1$, $\alpha \in N(g)\backslash N(f)$. For those α's we get by induction, if $k \geq 1$

$$| F_{k+l}(\alpha) | \leq \frac{1}{2^l} | F_k(\alpha) | + \frac{2^l - 1}{2^l},$$

hence, choosing any bound M for $\sup\limits_{\alpha \in N(g)} | F_k(\alpha) |$, then for $k \geq 1$ one gets

$$| F_{k+l}(\alpha) - sgn\ f(\alpha) | < \frac{1}{3}$$

provided $\frac{1}{2^l}M + 1 < \frac{4}{3}$, i.e. $\log_2(3M) < l$.

The final task remains to find a bound M. There are at least two ways. Set $F = F_k$, $k \geq 1$.

I) Consider $H(X) = Res_Y(g(Y), X - f(Y))$ then for $\beta \in \mathbb{R}$:

$$H(\beta) = 0 \Longleftrightarrow \beta = f(\alpha) \quad \text{for some} \quad \alpha \in N(g).$$

As is well-known the real roots of a polynomial h can be bounded in absolute value by $1 + ||h|| := 1 + \max\{|$ coefficients $\}$.

II) A less sophisticated bound can be obtained as follows:

$$| F(\alpha) |=| \sum_{0}^{N} a_i\alpha^i |\leq \sum | a_i | \cdot | \alpha |^i\leq ||F|| \cdot (1+||g||)^N$$

Clearly, F_k can be chosen with $deg\ F_k \leq n-1$ where $n = deg\ g$. Hence, $M \leq ||F_k|| \cdot (1+||g||)^{n-1}$.

We are going to summarize.

Proposition 3.2 *If $k \geq 1$, $\sup_{\alpha\in N(g)} | F_k(\alpha) |\leq M$ and the F_l are chosen with $deg\ F_l \leq n-1$ where $n = deg\ g$ then we have*

$$| F_{k+l}(\alpha) - sgn\ f(\alpha) |< \frac{1}{3} \quad \textit{for all} \quad \alpha \in N(g)$$

provided $l > log_2(3M)$. The bound M can be obtained as described above.

$\underline{2^{nd} \text{ case: } R = \mathbb{R}\{\epsilon_1\}\dots\{\epsilon_n\}}$

Basically, we follow the same approach as in the case $R = \mathbb{R}$. If $f \in R[T]$ there is a further polynomial denoted by $sgn\ f$ which satisfies

$$(sgn\ f)(\alpha) = sgn\ f(\alpha) \quad \text{for all roots} \quad \alpha \quad \text{of} \quad g \quad \text{in} \quad R, \quad \text{i.e.} \quad \alpha \in N(g).$$

As above, we try to find a polynomial h subject to

$$| h(\alpha) - (sgn\ f)(\alpha) |< \frac{1}{3} \quad \text{for all} \quad \alpha \in N(g).$$

To this end we want to use the Newton-method, i.e. we start by studying the sequence $f_{k+1} = \frac{1}{2}(f_k + \frac{1}{f_k}), k \geq 0, f_0 = f + (g(T))$, in $A_0 = {R[T]}\big/{(g(T))}$. However, in our present case, even if all f_k are units, the sequence (f_k) does not necessarily converge to $sgn\ f$ on $N(g)$. This failure is due to the fact that the order of R is non-Archimedean. We will remove this problem by modifying f into $\overline{f} \in R[T]$ where $sgn\ f(\alpha) = sgn\ \overline{f}(\alpha)$ for $\alpha \in N(g)$ and $\overline{f}$ allows a convergent Newton-sequence.

To prepare the construction of $\overline{f}$ we first study the behaviour of the mapping $a, b \longrightarrow a+\frac{1}{b}$ where $a, b \in R^*, ab > 0$. We will make use of the (Henselian) valuation v of R which arise from the recursive construction of R, is trivial on $\mathbb{R}$, has value group $\Gamma = \mathbb{Q} \times \dots \times \mathbb{Q}$, n-times, lexicographically ordered in ascending order of the factors, and residue field $\mathbb{R}$. In fact, every $a \in R^*$ has a unique presentation $a = \epsilon_1^{r_1} \dots \epsilon_n^{r_n} \cdot u, r_1, \dots, r_n \in \mathbb{Q}$, u a unit. We get $v(a) = (r_1, \dots, r_n) \in \Gamma$, and $a > 0$ iff the residue class of $u > 0$ in $\mathbb{R}$.

As usual, an element a is called infinitely small (resp. large) if $| a |< r$ (resp. $| a |> r$) for all $r \in \mathbb{R}, r > 0$. Equivalently, a is infinitely small (resp. large) if and only if $v(a) > 0$ (resp. $v(a) < 0$).

In particular, if a is any positive element with $v(a) \geq 0$ then $v(1 + a) = 0$. Using this one readily checks the following statements.

$(*)$ Assume $ab > 0$. Then

a) $0 \geq v(a) \geq v(b) \Rightarrow v(a + \frac{1}{b}) = v(a)$.

b) $v(a) \geq v(b) \geq 0 \Rightarrow v(a + \frac{1}{b}) = v(\frac{1}{b})$.

Consequently, for any $a \in R^*$ and setting $b = \frac{1}{2}(a + \frac{1}{a})$:

$(**)$

a) $v(a) < 0 \Rightarrow v(b) = v(a)$.

b) $v(a) > 0 \Rightarrow v(b) = -v(a) < 0$.

c) $v(a) = 0 \Rightarrow v(b) = 0$.

Now we consider the following Newton-sequence $x_{k+1} = \frac{1}{2}(x_k + \frac{1}{x_k}), x_0 = a$. Assume $v(x_0) = v(a) = 0$. Then $v(x_k) = 0$ for every $k \in \mathbb{N}$. We can write $x_k = \epsilon_k + m_k$ where $\epsilon_k \in \mathbb{R}^*, v(m_k) > 0$. Then $\epsilon_{k+1} = \frac{1}{2}(\epsilon_k + \frac{1}{\epsilon_k})$, hence (ϵ_k) converges in $\mathbb{R}$ to $sgn\ \epsilon_0 = sgn\ a$. This means we find for given $r \in \mathbb{R}^*_+$ $k(r) \in \mathbb{N}$ such that $| x_k - sgn(a) |< r$ for all $k \geq k(r)$.

Therefore, if an arbitrary element $a \in R^*$ is given and $sgn\,(a)$ should be computed via a Newton-sequence we first have to replace a by $\overline{a}$ subject to $v(\overline{a}) = 0, sgn\,(a) = sgn\,(\overline{a})$. Clearly, there is no problem if a is given explicitly. However, in our application we deal with the elements $f(\alpha), \alpha \in N(g)$ without knowing the roots $\alpha \in N(g)$, having only f, g at our disposal. As in the case $R = \mathbb{R}$ we consider $H(X) = Res_Y(g(Y), X - f(Y))$. We know that the roots of H in R are exactly the values $f(\alpha), \alpha \in N(g)$.

Therefore we are facing the following problem: given a polynomial $h \in R[T]$ design an algorithm constructing, on the input $a \in R^*$, an element $\overline{a}$ such that for all $\alpha \in N(h), \alpha \neq 0$ we get $v(\overline{\alpha}) = 0, sgn\ \alpha = sgn\ \overline{\alpha}$. The following algorithm is based on the observations listed in $(*)$ above.

1st step. Construct a list of elements $(x_i)_{i=1,\dots N}$ of R such that

1) $x_i > 0, i = 1, \dots, N$,

2) $v(x_1) > v(x_2) > \dots > v(x_N)$,

3) for every $\alpha \in N(h), \alpha \neq 0$ there is some x_i with $v(\alpha x_i) = 0$.

2^{nd} step. For each $a \in R^*$ compute $\overline{a}$ as the continued fraction

$$\overline{a} = [x_1 a, \ldots, x_N a] = x_1 a + \cfrac{1}{x_2 a + \cfrac{1}{x_3 a + \cfrac{1}{\ddots \cfrac{1}{x_{N-1} + \cfrac{1}{x_N a}}}}}$$

We first show that $\overline{\alpha}$ has the desired properties if $h(\alpha) = 0, \alpha \in R$. Obviously, $sgn\ \overline{a} = sgn\ a$ for every $a \in R$. Let $v(\alpha x_i) = 0$. Then $v(\alpha x_1) \geq \ldots \geq v(\alpha x_{i-1}) \geq v(\alpha x_i) = 0 \geq \ldots \geq v(\alpha x_N)$. Let $\overline{\alpha}_i = x_i \alpha + \cfrac{1}{x_{i+1}\alpha + \cfrac{1}{\ddots \cfrac{1}{x_N \alpha}}}$ denote the 'lower' part of $\overline{\alpha}$.

From $(*)$ we deduce $v(\overline{\alpha}_i) = 0$, $sgn\ \overline{\alpha}_i = sgn\ \alpha$. Again by using $(*)$ we derive $v(\overline{\alpha}) = 0$.

Thus, it remains to construct the list $(x_i)_{i=1,\ldots,N}$. Let $h = \sum\limits_{i=0}^{n} a_i X^i$, $\alpha \neq 0$ and $h(\alpha) = 0$. Since $h(\alpha) = 0$ there exist $i < j$ such that $a_i a_j \neq 0$ and $v(a_i \alpha^i) = v(a_j \alpha^j)$, i.e. $v(\alpha) = \frac{1}{j-i}(v(\frac{a_i}{a_j}))$. Therefore, if we can produce a positive element $x_{ij} \in R^*$ with $v(x_{ij}) = -\frac{1}{j-i} v(a_i/a_j)$ for each pair (i,j) such that $0 \leq i < j \leq n$ and $a_i, a_j \neq 0$ then the list (x_i) is obtained from ordering the x_{ij}'s accordingly. We have $\frac{a_j}{a_i} = \epsilon_1^{r_1} \ldots \epsilon_n^{r_n} \cdot u$, u a unit. Then set $x_{ij} = \epsilon_1^{s_1} \ldots \epsilon_n^{s_n}$ where $s_k = \frac{1}{j-i} r_k$, $k = 1, \ldots, n$. These elements have the desired properties.

Next, we transfer this algorithm to the global setting of polynomial functions on $N(g)$. Let $f \in R[T]$ and assume first that f and g are relatively prime. We will produce a polynomial $\overline{f}$ such that $\overline{f}(\alpha) = \overline{(f(\alpha))}$ for every $\alpha \in N(g)$. As already remarked above the values $f(\alpha)$ are the zeros of $h = Res_Y(g(Y), X - f(Y))$ in R. Then construct the list (x_i) attached to h in the algorithm above. In $R(T)$ we calculate the continued fraction $[x_1 f, \ldots, x_N f] =: f^*$. Using the recursion formulae for denominators of continued fractions we see that the denominators arising during the computation are of the type $f^r \cdot H$, $r = 0, 1$, $H \in R[T]$ a polynomial without zeros in R. This implies $f^*(\alpha) = [x_1 f(\alpha), \ldots, x_N f(\alpha)] = \overline{(f(\alpha))}$ for all $\alpha \in N(g)$. Set $f^* = \frac{A}{B}$, $A, B \in R[T]$ and $\overline{g} = g/gcd(g, B)$ where we assume $B = f^r H$ as above. Then $N(g) = N(\overline{g})$ and $gcd(B, \overline{g}) = 1$ since g is squarefree. From a presentation $1 = BC + D\overline{g}$, $C, D \in R[T]$ we derive that $B(\alpha)C(\alpha) = 1$ for every $\alpha \in N(\overline{g}) = N(g)$. Hence, $f^*(\alpha) = A(\alpha)C(\alpha)$ and $\overline{f} := A \cdot C$ is one of the wanted polynomials. Clearly, this polynomial $\overline{f}$ can be further reduced modulo $\overline{g}$ without loosing the property we are interested in.

We now drop the assumption that f and g are relatively prime. Then pass to $f_0 = f$ and $\tilde{g} = g/gcd(f, g)$. Compute $\overline{f}_0$ relative to f and $\tilde{g}$ as above, i.e. $\overline{f}_0(\alpha) = \overline{(f(\alpha))}$ for every $\alpha \in N(g)$ satisfying $f(\alpha) \neq 0$. As in the case $R = \mathbb{R}$ we find a polynomial χ satisfying $\chi(\alpha) = 1$ if $g(\alpha) = 0$, $f(\alpha) \neq 0$ and $\chi(\alpha) = 0$ if $g(\alpha) = 0 = f(\alpha)$. Then set $\overline{f} = \chi \cdot \overline{f}_0$.

Thus, in both cases we have constructed $\overline{f} \in R[T]$ satisfying for every $\alpha \in N(g)$: $\overline{f}(\alpha) = 0$ if $f(\alpha) = 0$, $\overline{f}(\alpha) = \overline{(f(\alpha))}$ if $f(\alpha) \neq 0$. Now the Newton-method can be applied to $\overline{f}$ to construct a sequence (F_k) in $R[T]$ approximating $sgn\ \overline{f}$. Now, since $\overline{f}(\alpha) = 0$ if $f(\alpha) = 0$ and $sgn\ \overline{f}(\alpha) = sgn\ f(\alpha)$ otherwise, from (F_k) we obtain a polynomial $h \in R[T]$ satisfying $\mid h(\alpha) - sgn\ f(\alpha) \mid < \frac{1}{3}$ as desired. □

4 A remark on Quantifier Elimination

The results of the first section allow us to give a very short and condensed quantifier-free expression for:

(i) $\exists x : g(x) = 0 \wedge f_1(x) > 0 \wedge \ldots \wedge f_m(x) > 0$ where $g \neq 0$;

(ii) $\exists x : f_1(x) > 0 \wedge \ldots \wedge f_m(x) > 0$.

By some well known arguments of model theory this can be extended to a proof of Quantifier Elimination in the theory of real closed fields. Let R be a real closed field and $Z \subset R$ a subring. If $g = a_nX^n + a_{n-1}X^{n-1} + \ldots + a_0, f_1, \ldots, f_m \in Z[X], a_n \neq 0$, we can use Prop. 2.2 to reformulate (i) as follows.

Let $A := R[X] \big/ (g(X))$ and x the canonical image of X in A.

Then (i) becomes:

$$(*) \quad \bigvee_{k=1}^{n} sgn(tr_{A|R})_* < \Pi f_i(x) > \otimes \ll f_1(x), \ldots, f_n(x) \gg = k.$$

This transfer decomposes orthogonally into transfers of the type

$$(tr_{A|R})_* < h(x) >$$

where h is expressible as $h = f_1^{\epsilon_1} \cdot \ldots \cdot f_m^{\epsilon_m}$, for $\epsilon_1, \ldots, \epsilon_m \in \{1, 2\}$. We can restrict ourselves to this case.

We are now going to sketch that $(*)$ is equivalent to a disjunction of quantifier-free expressions, polynomial over $\mathbf{Z}[1/a_n]$, in the coefficients of $g, f_1, \ldots, f_m$.

We note that $1, x, \ldots, x^{n-1}$ is a linear basis of the R-vectorspace A. With respect to this basis we get a matrix-presentation B for $(tr_{A|R})_* < h(x) >$ with entries

$$b_{ij} = (tr_{A|R})_*(h(x)x^ix^j) \quad i, j = 0, \ldots, n-1.$$

Definition 4.1 *The polynomials $M_j \in \mathbf{Z}[Z_1, \ldots, Z_j]$ defined by the recursion formula*

$$M_j + \sum_{i=1}^{j-1} Z_iM_{j-i} + jZ_j = 0, j \in \mathbb{N}_+,$$

are called Waring-polynomials.

We need to use the following classical result.

Lemma 4.2 *For monic g we have*

$$tr_{A|R}(x^j) = M_j(a_{n-1},\dots,a_0,0,\dots)\ j \in \mathbb{N}_+$$

$$tr_{A|R}(x^0) = n$$

Hence, if $h = c_k X^k + \dots + c_0$ we can express the (i,j)-th entry of the matrix B, setting $M_0 := n$ as:

$$\begin{aligned} tr_{A|R}(h(x)x^i x^j) &= \sum_{l=0}^{k} c_l \cdot tr_{A|R}(x^{l+i+j}) \\ &= \sum_{l=0}^{k} c_l\ M_{l+i+j}(\frac{a_{n-1}}{a_n},\dots,\frac{a_0}{a_n},0,\dots). \end{aligned}$$

There exist ways of expressing the signature of the symmetric matrix B by applying Descartes rule of signs to the characteristic polynomial χ_B of B. But these are not the most effective methods.

If $\chi_B = d_n X^n + d_{n-1} X^{n-1} + \dots + d_0$ let N_+ and N_- be the number of sign changes in the sequences

$$d_0,\dots,d_n$$

$$d_0,-d_1,\dots,(-1)^n d_n$$

respectively. Then Descartes rule implies that the signature of B is equal to the integer

$$N_+ - N_-.$$

Thus, the case i) is settled.

To deal with (ii) we use an idea of [KK]. Let $F := \prod_{i=1}^{m} f_i$; then we see that the polynomial

$$g = (1+F^2)^2(\frac{F}{1+F^2})' = F'(1-F^2)$$

has a root between any two adjacent roots of F, and also has a root in each of the intervals $(-\infty,\alpha_1)$, (α_2,∞), where α_1,α_2 are the smallest and the largest root, respectively.

Using those roots of g as testing points, we can replace ii) equivalently by:

$$\exists x : g(x) = 0 \wedge f_1(x) > 0 \wedge \dots \wedge f_m(x) > 0$$

Thus, we have arrived at the first case.

References

[Ba] Baeza, R.: Quadratic Forms over Semilocal Rings, Lect. Notes Math. 655, Springer (1978).

[Be1] Becker, E.: On the real spectrum of a ring and its application to semi-algebraic geometry, Bull. AMS 15 (1986), 19–60.

[Be2] Becker, E.: Sums of squares and quadratic forms in real algebraic geometry, Cahiers du Seminaire d'Histoire des Mathematique, 2^e Serie, Volume 1, Univ. P. et M. Curie (1991).

[BCR] Bochnak, J., Coste, M. and Roy, M.-F.: Géométrie Algébrique Reélle, Ergebnisse der Mathematik und ihrer Grenzgebiete (3. Folge) 12, Springer, Berlin Heidelberg New York (1987).

[BKR] Ben-Or, M., Kozen, D. and Reif, J.: The Complexity of Elementary Algebra and Geometry, J. Comp. System Sciences 32 (1986), 251-264.

[B–R] Benedetti, R. and Risler, J. J.: Real algebraic and semi-algebraic sets, Hermann, Éditeurs des sciences et de arts Paris (1990).

[Bu] Buchberger, B.: Gröbner bases: An algorithmic method in polynomial ideal theory, Multidimensional System Theory, Ed.: N. K. Bose, D. Reidel publishing company, Dordrecht, Boston Lancaster (1985), 184–232.

[C–E] Cartan, H. and Eilenberg, S.: Homological Algebra, Princeton University Press, Princeton 1956.

[C–R] Coste, M. and Roy, M.-F.: La topologie du spectre réel, Contemp. Math. 8 (1982), 27-59.

[DeM] De Meyer, F. and Ingraham, E.: Separable Algebras over Commutative Rings, Springer, Berlin Heidelberg New York (1971).

[E–L] Eisenbud, D. and Levine, H. I.: An algebraic formula for the degree of a C^∞ map germ, Annals of Math. 106 (1977), 19–44.

[G] Gantmacher, F. R.: Matrizentheorie, Springer, Berlin Heidelberg New York (1986).

[Gi–Mo] Gianni, P. and Mora, T.: Algebraic Solution of Systems of Polynomial Equations using Gröbner Bases, in: Lect. Notes Comp. Science, 356 (1987), 247-257.

[GLRR] Gonzalez, L., Lombardi, H., Recio, T. and Roy, M.-F.: Sturm–Habicht Sequence, Proc. of ISSAC–89 (1989), 136–146.

[GiTrZ] Gianni, P., Trager, B., Zacharias, G.: Gröbner bases and primary decomposition of polynomial ideals, J. Symb. Comp. 6 (1988), 149–167.

[I] Iohvidov, I. S.: Hankel and Toeplitz matrices and forms, Birkhäuser, Boston Basel Stuttgart (1982).

[K1] Knebusch, M.: On the uniqueness of real closures and the existence of real places, Comment. Math. Helv. 47 (1972), 657–673.

[K2] Knebusch, M.: Real closures of commutative rings I, J. reine angew. Math. 274/275 (1975), 61–80.

[K3] Knebusch, M., Rosenberg, A. and Ware, R.: Signatures on Frobenius extensions, Number theory and algebra, Academic Press, New York San Francisco London (1977), 167–186.

[K4] Knebusch, M.: Signaturen, reelle Stellen und reduzierte quadratische Formen, Jahresbericht Deutsche Math.-Verein 82 (1980), 109–127.

[K5] Knebusch, M.: An invitation to real spectra, in: Quadratic and hermitian forms, CMS Conf. Proc. 4 (1984), 51–105.

[KK] Korkina, E. I. and Kushnirenko, A. G.: Another proof of the Tarski-Seidenberg theorem, Sibirskii Mathematicheskii Zhurnal, Vol. 26, No. 5 (1985), 94–98, Sep.-Oct.

[K-S] Knebusch, M. and Scheiderer, C.: Einführung in die reelle Algebra, Vieweg, Braunschweig und Wiesbaden (1989).

[Kr-N] Krein, M. G. and Naimark, M. A.: The method of symmetric and hermitian forms in the theory of the separation of the roots of algebraic equations, Kharkov (1936) (in Russian) Engl. Transl. in: Lin. Multilin. Algebra 10 (1981), 265–308.

[M] Mahé, L.: Signatures et composantes connexes, Math. Ann. 260 (1982), 191–210.

[P1] Pedersen, P.: Generalizing Sturm's theorem to N dimensions, Technical Report, April 1990, New York University, Dept. of Computer Science, Courant Institute of Mathematical Science, New York.

[P2] Pedersen, P.: Counting Real Zeros, Technical Report, Nov. 1990, New York University, Dept. of Computer Science, Courant Institute of Mathematical Science, New York.

[P3] Pedersen, P., Roy, M.-F. and Szpirglas, A.: Counting real zeros in the multivariate case, to appear in Proc. of MEGA '92.

[S] Scheiderer, C.: Stability index of real varieties, Invent. Math. 97 (1989), 467–483.

[SchSt] Scheja, G. and Storch, U.: Lehrbuch der Algebra, B. G. Teubner, Stuttgart (1988), Band 2.

[Se] Seidenberg, A.: Constructions in Algebra, Trans. AMS 197 (1974), 273-313.