Sums of squares of linear forms: the quaternions approach

José F. Fernando, Jesús M. Ruiz, Claus Scheiderer

July 18, 2006

Abstract

Let $A = k[y]$ be the polynomial ring in one single variable y over a field k. We discuss the number of squares needed to represent sums of squares of linear forms with coefficients in the ring A. We use quaternions to obtain bounds when the Pythagoras number of A is ≤ 4. This provides bounds for the Pythagoras number of algebraic curves and algebroid surfaces over k.

1 Introduction

Let A be a commutative ring with 1. Following [ChDLR], we denote by $g_A(n)$ the smallest number ($\leq \infty$) of squares needed to represent any sum of squares of n-ary linear forms over A (that is, linear forms in $A[z_1, \ldots . z_n]$, for n variables $z_1, \ldots, z_n$). Of course, this is a variation of the *Pythagoras number* $p(B)$ of a ring B, which is the smallest number ($\leq \infty$) needed to represent any sum of squares of B. In fact, the invariant g_A can be used to bound the Pythagoras number of rings B that are finite modules over A. This goes back to Pfister who in the late '60s noted the following:

Remark 1.1 If $B \supset A$ is a finite A-module of rank n, then $p(B) \leq g_A(n)$.

Proof. Let $z_1, \ldots, z_n \in B$ generate B over A. Then a sum of squares $b \in B$ can be written as

$$b = \sum_{i=1}^{p} (a_{i1}z_1 + \cdots + a_{in}z_n)^2.$$

Then we look at the z_i's as variables, and at the expression above as a sum of squares of linear forms, so that for $g = g_A(n)$ we have:

$$\sum_{i=1}^{p} (a_{i1}z_1 + \cdots + a_{in}z_n)^2 = \sum_{j=1}^{g} (b_{j1}z_1 + \cdots + b_{jn}z_n)^2.$$

Finally, we look again at the z_j's as the generators, and we have expressed b as a sum of g squares in B. □

Of course, this remark is useful as soon as we can say something about g_A. This is the case if A is a field, as was immediately noticed by Pfister as follows:

Remark 1.2 Let $A = k$ be any field. Then $g_A(n) \leq p(A) \cdot n$ (*Pfister bound*).

Proof. Let $z_1, \ldots, z_n$ be variables, and consider a sum of squares of linear forms

$$Q = \sum_{i=1}^{p} (a_{i1}z_1 + \cdots + a_{in}z_n)^2.$$

By diagonalization over k, we get n linear forms

$$x_j = b_{j1}z_1 + \cdots + b_{jn}z_n,$$

and n coefficients $c_j \in k$ such that

$$Q = \sum_{j=1}^{n} c_j x_j^2.$$

But Q is obviously positive semidefinite in all orderings of k, hence the c_j's must be ≥ 0 in all orderings of k. Thus the c_j's are sums of squares in k, say of $p \leq p(A)$ squares:

$$c_j = c_{j1}^2 + \cdots + c_{jp}^2.$$

In the end, we get

$$Q = \sum_{j=1}^{n} \Big((c_{j1}x_j)^2 + \cdots + (c_{jp}x_j)^2 \Big),$$

and Q is a sum of $p \cdot n$ squares of linear forms, as wanted. □

This result can be used to deduce ([Rz]):

Remark 1.3 Let k be any field, and $A = k[[x]]$ the ring of formal power series in one single variable x. Then the Pfister bound holds for A.

Proof. There are some special cases. First, if k has characteristic 2, then $p(A[z_1, \ldots, z_n]) = 1$, by the identity $s^2 + t^2 = (s + t)^2$. Second, if 2 is a unit, but k is not real, then $p(A[z_1, \ldots, z_n]) \leq p(k) + 1$, as follows from the identity $t = \left(\frac{t+1}{2}\right)^2 - \left(\frac{t-1}{2}\right)^2$. Thus, we suppose k is *formally real*.

Let Q be a sum of squares of linear forms in the variables $z_1, \ldots, z_n$, with coefficients in the ring $A = k[[x]]$. By the previous Remark 1.2 for the field $k((x))$, we get

$$Q = \sum_{i=1}^{g} (a_{i1}z_1 + \cdots + a_{in}z_n)^2,$$

with $g \leq p(k((x))) \cdot n \leq p(A) \cdot n$. The problem is that the coefficients a_{ij} are in $k((x))$, and we want them in $A = k[[x]]$. But there is some power x^s such that

$$x^s Q = \sum_{i=1}^{g} (b_{i1}z_1 + \cdots + b_{in}z_n)^2,$$

with $b_{ij} \in k[[x]]$. Substituting $x = 0$ we get

$$0 = \sum_{i=1}^{g} (b_{i1}(0)z_1 + \cdots + b_{in}(0)z_n)^2.$$

As k is formally real, we deduce $b_{ij}(0) = 0$ for all i, j, that is, x divides all coefficients b_{ij}. Clearly by repeating this we get rid of the denominator x^s, and express Q as a sum of g squares of linear forms with coefficients in $A = k[[x]]$. □

It is interesting to describe the B's to which Remark 1.1 applies. For $A = k$, the typical B is a finitely generated algebra of Krull dimension 0, or in other words, a 0-*dimensional algebraic set* over k. For $A = k[[x]]$, B is a complete local ring whose residue field is a finite extension k, that is, an *algebroid curve* over k.

After these basic remarks, the first new result appears in [ChDLR]:

Proposition 1.4 *Let $A = R[y]$ be the polynomial ring in one single variable y over a real closed field R. Then $g_A(n) \leq 2n$.*

Since in this case $p(A) = 2$, this is the Pfister bound $g_A(n) \leq p(A) \cdot n$. The authors prove this by diagonalization, which over the ground ring $R[y]$ is far more involved than over a field. They stress how their proof needs the field R to be real closed. However, they conjecture that the same bound should hold for $A = k[y]$ under the mere assumption that $p(A) = 2$ (such a k is called *hereditarily Pythagorean,* see [Be]), More generally, they ask whether the finiteness of $p(A)$ implies that of $g_A(n)$, what can be more precisely stated as follows:

Is it true that $p(A) \leq p$ implies $g_A(n) \leq p\,n$ for $A = k[y]$, where k is a field?

This reformulation of the Pfister bound has some advantages. In fact, it is known from some results on the τ-invariant ([Sch]), that if $p(k[y]) \leq p = 2^s$, then $p(k((x))[y]) \leq p$. Hence, if we have the Pfister bound for such a p, then we have it for $A = k((x))[y]$, then for $A = k[[x]][y]$ (clearing the denominator x as in 1.1), and finally for $A = k[[x, y]]$ (by a suitable use of M. Artin's Approximation, see [Fe]). The most interesting rings here are the one we started with, $A = k[y]$, and the latter, $A = k[[x, y]]$. For these two, the typical B's in 1.1 are *algebraic curves* over k, and *algebroid surfaces* over k. We have so increased in one the dimension of the cases covered before. Of course, one cannot expect much more, as it is know that under quite mild assumptions, $p(B) = \infty$ if dimension is 3 or bigger ([FRS1]).

After this formulation of the problem, we know only of the progress in [Fe], which includes the Pfister bound for $A = k[y]$, where $k = \mathbb{R}(\{x\})$. This field is indeed hereditarily pythagorean, and in fact, most close to being real closed. The method is again a kind of diagonalization over A. The paper [Fe] deals with convergent power series over the reals because it is focused on the Pythagoras numbers of real analytic surface germs, rather than on algebroid surfaces over an arbitrary real closed field R. It would be possible to review the proofs there and extend them to $k = R((x))$.

The purpose of this note is to explore further the diagonalization technique, and prove the Pfister bound for $p \leq 4$, namely:

Theorem *Let $A = k[y]$ be the polynomial ring in one single variable y over a field k, and suppose that $p(A) \leq 2$ (resp. 4). Then $g_A(n) \leq 2n$ (resp. $4n$).*

As in [Fe], our main source of inspiration is [Dj], and quaternions play a central part in the game. One then thinks of octonions to reach the bound $8n$ when $p(A) \leq 8$, but there non-associativity stands in the way. Thus, to drop any assumption on $p(A)$ a different view must be taken. In fact, turning back to the ideas in [CLR], we have obtained in [FRS2] the following bound:

$$g_A(n) \leq 2n\,\tau(k),$$

where $\tau(k)$ is the τ-invariant mentioned earlier. Let us only say here that the Theorem above corresponds to the case $\tau(k) = 1$ (resp. 2).

The paper is organized as follows. In Section 2, we review some standard facts on quaternions and semidefinite matrices, mainly to fix notations. In Section 3 we state the diagonalization result, which is the core of the matter, and deduce from it the theorem above for $p = 4$; we also explain how it follows for $p = 2$, which is in fact easier. Section 4 contains the first half of the diagonalization theorem, namely the part that runs over the ring $k[y]$ until quaternions enter the scene. The proof of diagonalization is completed in Sections 5 and 6. It consists of an algorithm to make *dominant* certain matrices of quaternions.

The authors thank Prof. M. Coste for some useful discussions that lead to a more elegant formulation of the dominance algorithm.

2 Quaternions and semidefinite matrices

Let k be any field, and $A = k[y]$ the ring of polynomials in one single variable y with coefficients in k. We will consider *quaternions over k,* which are defined using three imaginary units u_1, u_2, u_3, that multiply as follows

- $u_i^2 = -1$,
- $u_1u_2 = u_3$, $u_2u_3 = u_1$, $u_3u_1 = u_2$, and
- $u_iu_j = -u_ju_i$ for $i < j$

Then, a quaternion is an element $a = (a_0, a_1, a_2, a_3) \in k^4$, which is costumarily denoted $a = a_0 + a_1u_1 + a_2u_2 + a_3u_3$. These quaternions form a k-algebra K with product defined through the rules ($\bullet$) above. This K is a skew field. The fact that K is not commutative is the source of many technical complications, but often we can use that at least the elements $a \in k$ commute with any other quaternion. Quaternions have a so-called *conjugation:*

$$a = a_0 + a_1u_1 + a_2u_2 + a_3u_3 \;\mapsto\; \overline{a} = a_0 - a_1u_1 - a_2u_2 - a_3u_3.$$

Clearly, k is the fixed part of this involution. One checks straighforwardly the two main properties

(i) $\overline{ab} = \overline{b}\overline{a}$, and

(ii) $a\overline{a} = \overline{a}a = a_0^2 + a_1^2 + a_2^2 + a_3^2$.

This is the well-known connection with sums of squares, which provides the classical formula for the product of two sums of four squares:

$$\begin{aligned}(a_0^2 + a_1^2 + a_2^2 + a_3^2)(b_0^2 + b_1^2 + b_2^2 + b_3^2) \\ &= (\overline{a}a)(\overline{b}b) = \overline{b}(\overline{a}a)b = (\overline{b}\overline{a})(ab) = \overline{c}c \\ &= c_0^2 + c_1^2 + c_2^2 + c_3^2,\end{aligned}$$

where $c = ab$ (note how we have used associativity, and that $\overline{a}a \in k$ conmutes with any other quaternion).

We now consider the ring of polynomials $K[y]$. Note here that the definition of a polynomial ring over a non-commutative ring assumes that the variable y commutes with every coefficient. This makes substitutions a delicate matter, but we will not need to discuss that here. Clearly, $K[y]$ is also the ring of *quaternions over* $k[y]$, consequently, we also call quaternions the elements of $K[y]$. Any element $a(y) \in K[y]$ can be written as

$$a(y) = a_0(y) + a_1(y)u_1 + a_2(y)u_2 + a_3(y)u_3, \quad a_i(y) \in k[y].$$

Thus we have the degree

$$\deg(a(y)) = \max_i \deg(a_i(y)),$$

and this notion behaves as usual for sums and products. Furthermore, in our situation, we can use Euclidean division to divide *component-wise* every polynomial quaternion in $K[y]$ by a polynomial in $k[y]$.

In the sequel we will use matrices with entries in $K[y]$. We will call them *q-matrices,* and reserve the term *matrices* for those with entries in $k[y]$. Diagonal q-matrices are denoted by $\langle a_1, \ldots, a_n\rangle$; in particular $I = \langle 1, \ldots, 1\rangle$. Since $K[y]$ is not commutative, operations with q-matrices, although defined as usual, must be computed with some care. In particular, there is no handy notion of determinant of a q-matrix, and we will avoid its use altogether. We say that a q-matrix α is *regular* if there is some other α' such that

$$\alpha\alpha' = \langle a_1, \ldots, a_n\rangle, \quad 0 \neq a_i \in K[y] \text{ for all } i.$$

If that is the case, multiplying on the right by the diagonal q-matrix

$$\langle \ldots, \overline{a}_i \prod_{j\neq i} a_j\overline{a}_j, \ldots\rangle$$

we obtain

$$\alpha\alpha'' = cI, \quad \text{where } c = \overline{a}a \in k[y],\ a = \prod_i a_i.$$

It is so clear that a *regular q-matrix simplifies by multiplication on the right*. The matrix α is *invertible* if there is another q-matrix α' such that

$$\alpha\alpha' = I.$$

It can be shown that then also $\alpha'\alpha = I$, but we will not need this.

Of great importance for us is the *transpose conjugate* α^* of a q-matrix α: if $\alpha = (\alpha_{ij})$, then $\alpha^* = (\overline{\alpha}_{ji})$. An straightforward computation shows that $(\alpha\beta)^* = \beta^*\alpha^*$. Then we say that α is *hermitean* if $\alpha = \alpha^*$. Note that this in particular implies that $\alpha_{ii} \in k[y]$ for all i. For matrices with entries in $k[y]$, hermitean means just symmetric.

Now let α be a hermitean q-matrix of order n. For every column $a = (a_i) \in K[y]^n$, the product $a^*\alpha a$ is in $K[y]$, and:

$$\overline{a^*\alpha a} = (a^*\alpha a)^* = a^*\alpha^* a = a^*\alpha a,$$

hence $a^*\alpha a \in k[y]$. We will say that α is positive semidefinite (psd or ≥ 0 in short) when $a^*\alpha a \geq 0$ in (every ordering of) the field $k(y)$, for every $a \in K[y]^n$.

For matrices with entries in $k[y]$ this can be defined without resource to quaternions. The reason is that writting $a \in K[y]^n$ as

$$a = a_0 + a_1u_1 + a_2u_2 + a_3u_3, \quad a_i \in k[y],$$

we obtain (after some computation):

$$a^*\alpha a = a_0^*\alpha a_0 + a_1^*\alpha a_1 + a_2^*\alpha a_2 + a_3^*\alpha a_3,$$

hence $a^*\alpha a \geq 0$ for all $a \in K[y]^n$ if and only if it holds for all $a_i \in k[y]$.

We conclude this review on q-matrices with the two properties that will be needed later:

Lemma 2.1 *(1) Let α, β and γ be q-matrices such that $\alpha = \gamma^*\beta\gamma$ is psd. If γ is regular, then β is psd.*

(2) Let α be a regular psd q-matrix. Then $\alpha_{ii} \neq 0$ for all i.

Proof. For (1), choose γ' with $\gamma\gamma' = c\langle 1, \ldots, 1\rangle$ with $0 \neq c \in k[y]$ (this γ' exists because γ is regular). Then $\gamma'^*\alpha\gamma'$ is psd (this is the converse of (1), which is immediate). But

$$\gamma'^*\alpha\gamma' = \gamma'^*\gamma^*\beta\gamma\gamma' = c^2\beta,$$

and since $0 \neq c \in k[y]$, it follows readily that β is also psd.

(2) Clearly, since α is regular, no row vanishes, so that $\alpha_{ij} \neq 0$ for some j. If $j = i$ we are done. Otherwise, suppose $\alpha_{ii} = 0$, and consider the q-matrix of order 2

$$\begin{pmatrix} \alpha_{ii} & \alpha_{ij} \\ \alpha_{ji} & \alpha_{jj} \end{pmatrix} = \begin{pmatrix} 0 & c \\ \overline{c} & b \end{pmatrix}, \quad \text{where } b \in k[y],\ c \in K[y].$$

Since α is psd, this matrix is psd too, hence

$$0 \leq (-b - \tfrac{1}{2}, c) \begin{pmatrix} 0 & c \\ \overline{c} & b \end{pmatrix} \begin{pmatrix} -b - \frac{1}{2} \\ \overline{c} \end{pmatrix} = -(b + \tfrac{1}{2})\overline{c}c.$$

But $b, \overline{c}c \in k[y]$ are both ≥ 0, hence we can only conclude that $0 = c = \alpha_{ij}$. Contradiction. □

3 The diagonalization theorem

Consider a field k with $p(k[y]) \leq 4$. The Pfister bound for $A = k[y]$ follows by diagonalization of psd matrices, but that diagonalization is performed over the quaternions. Namely:

Theorem 3.1 *Let α be a psd matrix (with entries in $k[y]$). Then there are a regular q-matrix γ and polynomials $a_1, \ldots, a_n \in k[y]$ such that*

$$\alpha = \gamma^* \langle a_1, \ldots, a_n \rangle \gamma.$$

The proof of this will be developed in the next three sections. Here, we deduce from it the announced Pfister bound:

Corollary 3.2 *For A as above, we have $g_A(n) \leq 4n$.*

Proof. Consider a sum of squares of quadratic forms

$$Q = \sum_{i=1}^{p} (a_{i1}z_1 + \cdots + a_{in}z_n)^2, \quad a_{ij} \in A = k[y].$$

Writting the variables as a column z, we have $Q = z^* \alpha z$, where α is a well defined psd matrix. Then diagonalization gives

$$\alpha = \gamma^* \langle a_1, \ldots, a_n \rangle \gamma.$$

Since the diagonal matrix must be psd (Lemma 2.1(1)), each a_i is ≥ 0 in (all orderings of) $k(y)$, hence a sum of four squares in $k[y]$, so that

$$a_i = \overline{b}_i b_i, \quad \text{for some quaternion } b_i \in K[y].$$

Thus:

$$\langle a_1, \ldots, a_n \rangle = \beta^* \beta, \quad \text{where } \beta = \langle b_1, \ldots, b_n \rangle,$$

and

$$\alpha = \gamma^* \beta^* \beta \gamma = \theta^* \theta, \quad \text{where } \theta = \beta\gamma;$$

finally, consider the four components of the entries $\theta_{ij} \in K[y]$:

$$\theta_{ij} = \theta_{ij}^{(0)} + \theta_{ij}^{(1)} u_1 + \theta_{ij}^{(2)} u_2 + \theta_{ij}^{(3)} u_3.$$

Now we are ready for the key computation:

$$\begin{aligned} Q &= z^* \alpha z = z^* \theta^* \theta z = (\theta z)^* (\theta z) \\ &= \sum_{i=1}^{n} \overline{(\theta_{i1} z_1 + \cdots + \theta_{in} z_n)} (\theta_{i1} z_1 + \cdots + \theta_{in} z_n) \\ &= \sum_{i=1}^{n} \sum_{\ell=0}^{3} (\theta_{i1}^{(\ell)} z_1 + \cdots + \theta_{in}^{(\ell)} z_n)^2. \end{aligned}$$

This latter expression is a sum of $4n$ squares of linear forms, and we are done. □

Remarks 3.3 (1) The above Pfister bound corresponds to the case $p(A) \leq p = 4$. For $p = 2$ the argument is the same, only that simpler. Indeed, for $p = 2$ quaternions are not needed: one works over the complexes $C = k[u_1]$, that is, both imaginary units u_2 and u_3 are forgotten. This has the advantage that C is commutative and the same computations become much easier.

(2) Of course, the next question is whether the case $p = 8$ could be treated using octonions. This will in particular include the field $k = \mathbb{Q}$ in our results. However, octonions are much worse than quaternions concerning computations, because they are *not associative*. We do not know how to overcome this difficulty.

4 Reduction of diagonalization to the invertible case

Here we develop the first part of the proof of diagonalization, which is a reduction to the invertible case.

Let α be a psd matrix (with entries in $k[y]$).

Step I. Diagonalization over $k[y]$.

Since $k[y]$ is a principal ideal domain, we know ([Hu, VII.2]) that there exist two invertible matrices u, v and a diagonal matrix

$$e = \langle e_1, \ldots, e_r, 0, \ldots, 0 \rangle,$$

such that $e_1 | e_2 | \ldots | e_r$ and $\alpha = uev$. Clearly, we can choose all $e_i \in k[y]$ to be monic.

Then we write

$$\alpha = \gamma^* e \pi \gamma, \quad \text{with } \gamma^* = u,\ \pi = v\gamma^{-1}.$$

Notice that the matrices γ and π are invertible. This implies that $\beta = e\pi$ is psd. Furthermore, since $\alpha = \alpha^*$ we see that $e\pi = \pi^* e$, so that

$$e_i \pi_{ij} = \pi_{ji} e_j.$$

Similarly, if $\pi\pi' = I$, then

$$e\pi' = (\pi'^*\pi^*)e\pi' = \pi'^*(\pi^* e)\pi' = \pi'^*(e\pi)\pi' = \pi'^* e(\pi\pi') = \pi'^* e,$$

so that:

$$e_i\pi'_{ij} = \pi'_{ji}e_j.$$

Next, splitting our matrices into boxes with $\widetilde{e} = \langle e_1, \dots, e_r\rangle$, the equation $e\pi = \pi^* e$ becomes:

$$\begin{pmatrix} \widetilde{e} & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} \widetilde{\pi} & \theta \\ \rho & \eta \end{pmatrix} = \begin{pmatrix} \widetilde{\pi}^* & \rho^* \\ \theta^* & \eta^* \end{pmatrix}\begin{pmatrix} \widetilde{e} & 0 \\ 0 & 0 \end{pmatrix}$$

which implies $\widetilde{e}\,\theta = 0$, hence $\theta = 0$. It follows that $\widetilde{\pi}$ is invertible, and we have an equation

$$\alpha = \gamma^*\begin{pmatrix} \widetilde{e}\,\widetilde{\pi} & 0 \\ 0 & 0 \end{pmatrix}\gamma.$$

Consequently, it is enough to prove the diagonalization result for $\widetilde{e}\,\widetilde{\pi}$. Thus, we can henceforth suppose $r = n$ and $\alpha = e\pi$. □

Step II. All $e_i \in k[y]$ are sums of four squares.

Since $p(k[y]) \le 4$, by a classical theorem of Cassels ([Ca]), it is enough to show that all e_i's are sums of squares in $k(y)$, or equivalently, that they are positive in all orderings of $k(y)$. We argue by way of contradiction. Suppose $e_\ell < 0$ in some ordering of $k(y)$, and all preceding e_i are positive in all orderings of $k(y)$. Then by the Artin-Lang Homomorphism Theorem ([BCR, 4.1.2]), the polynomial $e_\ell \in k[y]$ has some negative specialization in a real closure R of k, and since e_ℓ is monic, it certainly has also positive specializations. Thus, e_ℓ has in R some root ξ of odd multiplicity μ_ℓ. We denote μ_i the multiplicity of ξ as a root of e_i (which maybe zero). Since for $i \le j$ we have $e_i | e_j$, it is $\mu_i \le \mu_j$; on the other hand, since for $i < \ell$ the polynomial e_i is positive in all orderings of $k(y)$, μ_i must be even, hence $\mu_i < \mu_\ell$. We now consider the specialization $\pi(\xi)$ of the matrix π, and look for zero entries $\pi_{ij}(\xi)$.

1. Case $i = \ell = j$. Since $e\pi$ is psd, then $P = e_\ell\pi_{\ell\ell}$ is positive in every ordering of $k[y]$, hence ξ must be a root of even multiplicity of P. But its multiplicity μ_ℓ in e_ℓ is odd, hence $\pi_{\ell\ell}(\xi) = 0$.

2. Case $i < \ell \le j$. As $e_i\pi_{ij} = \pi_{ji}e_j$ and $\mu_i < \mu_j$, it must be $\pi_{ij}(\xi) = 0$.

3. Case $i = \ell < j$. The matrix
$$\begin{pmatrix} e_\ell\pi_{\ell\ell} & e_\ell\pi_{\ell j} \\ e_j\pi_{j\ell} & e_j\pi_{jj} \end{pmatrix}$$
is psd, because so is $e\pi$. Now, since $k[y]$ is commutative, we can look at its determinant, which must be ≥ 0 in all orderings of $k(y)$. But $e_j\pi_{j\ell} = \pi_{\ell j}e_\ell$ and $e_\ell | e_j$, so that
$$\det\begin{pmatrix} e_\ell\pi_{\ell\ell} & e_\ell\pi_{\ell j} \\ e_j\pi_{j\ell} & e_j\pi_{jj} \end{pmatrix} = \det\begin{pmatrix} e_\ell\pi_{\ell\ell} & e_\ell\pi_{\ell j} \\ e_\ell\pi_{\ell j} & e_j\pi_{jj} \end{pmatrix} = e_\ell^2\Big(\frac{e_j}{e_\ell}\pi_{\ell\ell}\pi_{jj} - \pi_{\ell j}^2\Big).$$

Since $e_\ell^2 > 0$ in all orderings, we deduce that

$$\Delta = \frac{e_j}{e_\ell}\pi_{\ell\ell}\pi_{jj} - \pi_{\ell j}^2 \geq 0$$

in all orderings. Thus all specializations of this element Δ must be ≥ 0, hence

$$0 \leq \Delta(\xi) = -\pi_{\ell j}(\xi)^2$$

(we already know that $\pi_{\ell\ell}(\xi) = 0$). We conclude that $\pi_{\ell j}(\xi) = 0$.

Consequently, with that many zero entries, the determinant of the matrix $\pi(\xi)$ must be zero. But $\pi(\xi)$ is a specialization of the *invertible* matrix π, hence it must be invertible too. This contradiction ends the proof. □

Step III. Factorization of e over the quaternions.

By the preceding step, we can write $e_1 = \overline{\varepsilon}_1\varepsilon_1$ for some $\varepsilon_1 \in K[y]$. Now, as $e_1|e_2$, the quotient $f_2 = e_2/e_1 \in k[y]$ is also a sum of four squares, so $f_2 = \overline{\varphi}_2\varphi_2$ with $\varphi_2 \in K[y]$. Setting $\varepsilon_2 = \varphi_2\varepsilon_1$ we get:

$$e_2 = (e_2/e_2)e_1 = \overline{\varphi}_2\varphi_2\overline{\varepsilon}_1\varepsilon_1 = \overline{\varepsilon}_1\overline{\varphi}_2\varphi_2\varepsilon_1 = \overline{\varphi_2\varepsilon_1}\varphi_2\varepsilon_1 = \overline{\varepsilon}_2\varepsilon_2.$$

Clearly, by repetition, we end up with

$$\begin{cases} e_i = \overline{\varepsilon}_i\varepsilon_i, & \varepsilon_i|\varepsilon_{i+1} \text{ in } K[y], \text{ and} \\ e = \varepsilon^*\varepsilon, & \text{where the } q\text{-matrix. } \varepsilon = \langle\varepsilon_1, \ldots, \varepsilon_n\rangle \text{ is regular.} \end{cases}$$

□

Step IV. There is an invertible q-matrix σ such that $e\pi = \varepsilon^*\sigma\varepsilon$.

As $e\pi = \varepsilon^*\varepsilon\pi$, we look for σ such that $\varepsilon\pi = \sigma\varepsilon$. Now, we have the regular matrix

$$\varepsilon' = \langle \ldots, \overline{\varepsilon}_j \prod_{\ell\neq j} \varepsilon_\ell\overline{\varepsilon}_\ell, \ldots\rangle$$

such that

$$\varepsilon\varepsilon' = cI, \quad \text{where } c = \prod_i e_i.$$

Consequently we need that

$$\varepsilon\pi\varepsilon' = \sigma\varepsilon\varepsilon' = c\sigma.$$

(and note that ε' simplifies in the first equality). Thus, we will conclude by showing that c divides all entries of the q-matrix $\varepsilon\pi\varepsilon' = (\theta_{ij})$. We have:

$$\theta_{ij} = \varepsilon_i\pi_{ij}\overline{\varepsilon}_j \prod_{\ell\neq j} \varepsilon_\ell\overline{\varepsilon}_\ell = \varepsilon_i\pi_{ij}\overline{\varepsilon}_j \prod_{\ell\neq j} e_\ell,$$

hence we must check that $e_j|\varepsilon_i\pi_{ij}\overline{\varepsilon}_j$. We distinguish two cases:

1. If $j \leq i$, write $\varepsilon_i \pi_{ij} \overline{\varepsilon}_j = \pi_{ij} \varepsilon_i \overline{\varepsilon}_j = \pi_{ij}(\varepsilon_i/\varepsilon_j)\varepsilon_j \overline{\varepsilon}_j = \pi_{ij}(\varepsilon_i/\varepsilon_j)e_j$.

2. If $j > i$, write $\varepsilon_i \pi_{ij} \overline{\varepsilon}_j = \varepsilon_i \pi_{ij} \overline{\varepsilon}_i \overline{(\varepsilon_j/\varepsilon_i)} = \varepsilon_i \overline{\varepsilon}_i \pi_{ij} \overline{(\varepsilon_j/\varepsilon_i)} = e_i \pi_{ij} \overline{(\varepsilon_j/\varepsilon_i)} = e_j \pi_{ji} \overline{(\varepsilon_j/\varepsilon_i)}$ (recall that $e_i \pi_{ij} = e_j \pi_{ji}$).

In both cases we get an element divisible by e_j, as wanted.

It remains to see that σ is invertible, that is, to find σ' such that $\sigma\sigma' = I$. But suppose we find σ' such that $\varepsilon\pi' = \sigma'\varepsilon$. Then

$$\sigma\sigma'\varepsilon = \sigma\varepsilon\pi' = \varepsilon\pi\pi' = \varepsilon,$$

and simplifying ε we are done. Once this said, one finds σ' alike σ. □

After Steps I-IV it suffices to prove:

Proposition 4.1 *Let σ be an invertible psd q-matrix, then there exists an invertible q-matrix λ such that $\lambda^*\sigma\lambda$ is diagonal.*

Indeed, Step I reduces the problem to the product $e\pi \geq 0$, and Steps II-IV provide a factorization $e\pi = \varepsilon^*\sigma\varepsilon$, and $\sigma \geq 0$ because ε is regular. Now if λ is invertible and $\theta = \lambda^*\sigma\lambda$ diagonal, we have $\lambda\lambda' = I$ for a suitable λ', and

$$e\pi = \varepsilon^*\lambda'^*\theta\lambda'\varepsilon = \gamma^*\theta\gamma,$$

where $\gamma = \lambda'\varepsilon$ is regular.

5 Dominant matrices

For the proof of Proposition 4.1, we will resource to a notion that guarantees in a very strong sense that a q-matrix is diagonal:

Definition 5.1 *We say that a q-matrix τ of order n is* dominant *if* $\deg(\tau_{ii}) > \deg(\tau_{ij})$ *for all $j \neq i$, and* $\deg(\tau_{11}) \leq \cdots \leq \deg(\tau_{nn})$.

Note that if τ is hermitean (or symmetric), it is enough to check $\deg(\tau_{ii}) > \deg(\tau_{ij})$ for $j > i$.

The key fact is that a dominant invertible psd q-matrix τ is diagonal. Suppose first $\deg(\tau_{11}) = 0$. Then, by dominance, the other elements in the first row vanish, and since τ is hermitean, also the others in the first column. Consequently, by induction, it is enough to show that indeed $\deg(\tau_{11}) = 0$. We argue by way of contradiction, assuming that all entries in the diagonal have degree > 0. Since τ is invertible, there is τ' such that $\tau\tau' = I$, that is

$$\sum_{\ell=1}^{n} \tau_{i\ell}\tau'_{\ell j} = \delta_{ij}.$$

For $j = 1$ we get

$$\begin{aligned}\deg(\tau_{ii}) + \deg(\tau'_{i1}) &\leq \max_{\ell\neq i}\{\deg(\tau_{i\ell}) + \deg(\tau'_{\ell 1})\}\\ &< \max_{\ell\neq i}\{\deg(\tau_{ii}) + \deg(\tau'_{\ell 1})\} \quad \text{(dominance)}\\ &= \deg(\tau_{ii}) + \max_{\ell\neq i}\{\deg(\tau'_{\ell 1})\}.\end{aligned}$$

Hence,

$$\deg(\tau'_{i1}) < \max_{\ell\neq i}\{\deg(\tau'_{\ell 1})\}, \quad \text{for all } i.$$

This is impossible.

We have thus shown that τ is diagonal. But even more, we have shown that τ has constant coefficients! Anyway, Theorem 4.1 reduces to:

Proposition 5.2 *Let σ be an invertible psd q-matrix, then there exists an invertible q-matrix λ such that $\tau = \lambda^*\sigma\lambda$ is dominant.*

Indeed, such a dominant τ is invertible and psd as σ is.

6 Diagonalization by dominance

In this section we describe an algorithm to make dominant an invertible psd q-matrix σ, as stated in Proposition 5.2. For this we will describe two operations of the allowed type $\lambda^*\sigma\lambda$.

(6.1) Permutation of columns and rows. We denote by $I_{[i,j]}$ the invertible matrix obtained from the identity by interchanging the i-th and j-th columns, $i < j$. If α is any q-matrix, then $\alpha I_{[i,j]}$ is obtained from α in the same way: by interchanging the i-th and j-th columns. We can also describe $I^*_{[i,j]} = I_{[i,j]}$ as obtained from the identity by interchanging the i-th and j-th rows, and then, for any q matrix β, the product $I^*_{[i,j]}\beta$ results from β by interchanging the i-th and j-th rows. Putting all together, given a q-matrix σ, for the product $\nu = I^*_{[i,j]}\sigma I_{[i,j]}$ we have:

$$\begin{cases}\nu_{\ell i} = \sigma_{\ell i} & \text{for } \ell < i, \text{ and}\\ \nu_{ii} = \sigma_{jj}.\end{cases}$$

We will denote $\nu = \sigma[i,j]$. □

Next, we need some more notation:

Definition 6.2 *A q-matrix σ of order n is called* r-dominant *if its* initial minor $(\sigma_{ij})_{i,j\leq r}$ *is dominant. For $r = 1$ this just means $\sigma_{11} \neq 0$, and for $r = n$ we recover dominance. If $r < n$ we consider the n-tuple*

$$w_r(\sigma) = (\deg(\sigma_{11}), \ldots, \deg(\sigma_{rr}), \infty, \ldots, \infty),$$

and the integer

$$d_r(\sigma) = \max_{i\le r}\{\deg(\sigma_{i\ r+1}) - \deg(\sigma_{ii})\}.$$

Using this we describe now a second operation

(6.3) Division procedure. Let σ be an r-dominant q-matrix of order n, with $r < n$ and $d = d_r(\sigma) \ge 0$. Then:

1. Define $q_1, \dots, q_r \in K[y]$ by induction through Euclidean division:

$$\begin{cases} \sigma_{1\,r+1} = \sigma_{11}q_1 + \rho_1, & \deg(\rho_1) < \deg(\sigma_{11}), \\ \sigma_{i\ r+1} - \sum_{j<i}\sigma_{ij}q_j = \sigma_{ii}q_i + \rho_i\,, & \deg(\rho_i) < \deg(\sigma_{ii}), \qquad \text{for } i > 1 \end{cases}$$

2. Consider the invertible matrix λ which coincides with the identity except for the entries

$$\lambda_{i\ r+1} = -q_i, \quad 1 \le i \le r,$$

and multiply $\nu = \lambda^*\sigma\lambda$, to get

$$\begin{cases} \nu_{ij} = \sigma_{ij}, & \text{for } i,j \le r, \\ \nu_{i\ r+1} = \sigma_{i\ r+1} - \sum_{j\le r}\sigma_{ij}q_j = \rho_i - \sum_{i<j\le r}\sigma_{ij}q_j, & \text{for } i \le r \end{cases}$$

(in particular, $\nu_{r\,r+1} = \rho_r$). We claim that $d_r(\nu) < d$.

Indeed, firstly note that $\deg(q_1) = \deg(\sigma_{1r+1}) - \deg(\sigma_{11}) \le d$, and by induction:

$$\begin{aligned} \deg(q_i) &\le \max_{j<i\le r}\{\deg(\sigma_{i\ r+1}), \deg(\sigma_{ij}) + \deg(q_j)\} - \deg(\sigma_{ii}) \\ &\le \max_{j<i\le r}\{\deg(\sigma_{i\ r+1}) - \deg(\sigma_{ii}), \deg(\sigma_{ij}) - \deg(\sigma_{ii}) + \deg(q_j)\} \\ &\le \max\{d, 0+d\} = d \end{aligned}$$

(the zero comes from r-dominance). Next, we have

$$\deg(\nu_{r\,r+1}) - \deg(\nu_{rr}) = \deg(\rho_r) - \deg(\sigma_{rr}) < 0 \le d,$$

and by descending induction:

$$\begin{aligned} \deg(\nu_{i\ r+1}) - \deg(\nu_{ii}) &\le \max_{i<j\le r}\{\deg(\rho_i), \deg(\sigma_{ij}) + \deg(q_j)\} - \deg(\sigma_{ii}) \\ &\le \max_{i<j\le r}\{\deg(\rho_i) - \deg(\sigma_{ii}), \deg(\sigma_{ij}) - \deg(\sigma_{ii}) + \deg(q_j)\} \\ &< \max_{i<j\le r}\{0, 0 + \deg(q_j)\} \le d \end{aligned}$$

(the first zero comes from Euclidean division, the second from r-dominance).

Once the claim is proved, it is clear that by repetition we get a matrix ν such that

$$\begin{cases} \nu_{ij} = \sigma_{ij}, & \text{for } i,j \le r, \text{ and} \\ \deg(\nu_{i\ r+1}) < \deg(\nu_{ii}), & \text{for } i \le r. \end{cases}$$

We will denote $\nu = \sigma[r]$. $\square$

Once these two procedures are defined, we can describe the

(6.4) Algorithm for dominance. Let σ be an invertible psd q-matrix of order n. Then $\sigma_{11} \neq 0$ (Lemma 2.1(2)), hence σ is r-dominant for some maximum $r \geq 1$. If $r = n$, then $\tau = \sigma$ is dominant. Otherwise, consider $\nu = \sigma[r]$. If $\deg(\nu_{r+1\ r+1}) \geq \deg(\nu_{rr})$, then ν is s-dominant with $s > r$, and we set $\tau = \nu$. If $\deg(\nu_{r+1\ r+1}) < \deg(\nu_{rr})$, let t be the smallest index such that $\deg(\nu_{r+1\ r+1}) < \deg(\nu_{tt})$, and set $\tau = \nu[t, r+1]$. This q-matrix is s-dominant with maximum $s \geq t$ (maybe $s < r$) and $\deg(\tau_{tt}) < \deg(\nu_{tt}) = \deg(\sigma_{tt})$.

This will eventually produce a dominant q-matrix τ. Indeed, by construction

$$w_r(\sigma) = (a_1, \ldots, a_r, \infty, \ldots, \infty) > (b_1, \ldots, b_s, \infty, \ldots, \infty) = w_s(\tau)$$

in the lexicographic ordering, because: either (i) the dominance order increases, and some infinite entry becomes finite (case $\tau = \nu$), or (ii) the degree of some entry in the dominant initial minor decreases (case $\tau = \nu[t, r+1]$). Since the lexicographic ordering has no infinite descending chains, the algorithm stops after finitely many repetitions.

Thus we have proved Proposition 5.2, hence Proposition 4.1, hence the diagonalization theorem.

References

[Be] E. Becker: Hereditarily-Pythagorean fields and orderings of higher level. *Monografías de Matemática [Mathematical Monographs]*, **29.** Rio de Janeiro: Instituto de Matemática Pura e Aplicada, 1978.

[BCR] J. Bochnak, M. Coste, M.F. Roy: Real Algebraic Geometry. *Ergeb. Math.* **36.** Berlin Heidelberg New York: Springer-Verlag, 1998.

[Ca] J.W.S. Cassels: On the representation of rational functions as sums of squares. *Acta Arith.* **9** (1964) 79–82.

[ChDLR] M.D. Choi, Z.D. Dai, T.Y. Lam, B. Reznick: The Pythagoras number of some affine algebras and local algebras. *J. reine Angew. Math.* **336** (1982) 45–82.

[CLR] M.D. Choi, T.Y. Lam, B. Reznick: Real zeros of positive semidefinite forms. I. Math. Z. **171**, 1–26 (1980).

[Dj] D.Z. Djoković: Hermitean matrices over polynomial rings. *J. Algebra* **43** (1976) 359–374.

[Fe] J.F. Fernando: On the Pythagoras numbers of real analytic rings. *J. Algebra* **356** (2004) 2663–2684.

[FRS1] J.F. Fernando, J.M. Ruiz, C. Scheiderer: Sums of squares in real rings. *Trans. AMS* **243** (2001) 321–338.

[FRS2] J.F. Fernando, J.M. Ruiz, C. Scheiderer: Sums of squares of linear forms. To appear in *Math. Research Letters.*

[Hu] T.W. Hungerford: Algebra. *Graduate Text in Math.* **73.** Berlin Heidelberg New York: Springer Verlag, 1974.

[Rz] J.M. Ruiz: On pythagorean real algebroid curves. *Rocky Mountain J. Math.* **14** (1984) 899–901.

[Sch] C. Scheiderer: On sums of squares in local rings. *J. reine angew. Math.* **540** (2001), 205–227.

José F. Fernando
Depto. Matemáticas
Facultad de Ciencias
Univ. Autónoma de Madrid
Cantoblanco, 28049 Madrid, Spain
josefrancisco.fernando@uam.es

Jesús M. Ruiz
Depto. Geometría y Topología
Facultad de Matemáticas
Univ. Complutense de Madrid
28040 Madrid, Spain
jesusr@mat.ucm.es

Claus Scheiderer
Fach. Mathematik und Statistik
Fach D 203
Universität Konstanz
78457 Konstanz, Germany
claus.scheiderer@uni-konstanz.de