Proceedings of Symposia in Pure Mathematics
Volume 58.2 (1995)

# Sums of Squares of Real Polynomials

M. D. CHOI, T. Y. LAM, AND B. REZNICK

## §1. Introduction

Given that an element $a$ in a ring $A$ is a sum of squares in $A$, say $a = \sum_{i=1}^{t} a_i^2$, it is natural to ask for the smallest $t$ (called the *length* of $a$ in $A$) for which such an expression is possible. For any subset $S$ of $A$, the *pythagoras number* $P(S)$ of $S$ is defined to be $\sup\{\text{length}(a)\}$, where $a$ ranges over all elements of $S$ which are sums of squares in $A$. The computation of $P(S)$ is an interesting, but often difficult task.

For the field of rational functions $K_n = \mathbf{R}(x_1, \dots, x_n)$ over the real field $\mathbf{R}$, Pfister **[Pf]** has shown that $P(K_n) \leq 2^n$. This upper bound is known to be sharp only for $n \leq 2$ **[CEP]**. As for lower bounds, one can deduce from **[CEP]** that, for $n \geq 2$, $P(K_n) \geq n+2$. The precise value of $P(K_n)$ for $n \geq 3$ remains unknown. On the other hand, Z. D. Dai and the present authors **[CDLR]** have shown that, for the polynomial ring $A_n = \mathbf{R}[x_1, \dots, x_n]$, $P(A_n) = \infty$ for $n \geq 2$ (while, of course, $P(A_1) = 2$).

The results mentioned above provides the backdrop of this work, in which we study the problem of computing the lengths of homogeneous polynomials (or *forms*) in $A_n$. In doing so, we shall not be interested in writing forms as sums of squares of rational functions; instead, we shall only be interested in writing forms as sums of squares of other polynomials (necessarily forms). Thus, throughout this paper, "*sums of squares*" shall always mean *sums of squares of polynomials* (or *forms*). We note in passing that any sum of squares is psd (= "positive semidefinite"), though in general a psd polynomial need not be a sum of squares **[Hi]**.

An $n$-ary $m$-ic shall mean a form $f = f(x_1, \dots, x_n) \in A_n$ of degree $m$. Let $F_{n,m}$ denote the $\mathbf{R}$-space of $n$-ary $m$-ics in $A_n$. Our main object of

1991 *Mathematics Subject Classification.* Primary 11E25, 11E76, 11E81, 12D15; Secondary 12E99, 14E99, 14N99, 52B50.

The research of the first author was supported in part by NSERC of Canada.
The research of the second and third authors was supported in part by NSF.
This paper is in final form and no version of it will be submitted for publication elsewhere.

study in this paper is the set of pythagoras numbers $P(n, m) := P(F_{n,m})$. Two well-known examples are: $P(2, m) = 2$ (for binary $m$-ics, $m > 0$) and $P(n, 2) = n$ (for $n$-ary quadratics). A remarkable theorem of Hilbert **[Hi]** says that $P(3, 4) = 3$ for ternary quartics.[1] No other values of $P(n, m)$ have appeared in the literature, except that R. M. Robinson **[Ro]** has shown that $P(n, m) \le \binom{n+m-1}{n-1}$. (In particular, $P(n, m)$ is finite, which is not *a priori* obvious).

In the present work, we develop a general method for studying sums of squares in the polynomial ring $A_n$. The gist of our method lies in the observation (§2) that, if a form $f$ is written, *in a specific way*, as a sum of squares, say $f = \sum h_i^2$, then we can associate to this expression a "Gram matrix" obtained from dot products of certain vectors arising from the coefficients of the $h_i$'s. By considering *all* expressions of $f$ as a sum of squares, we get a family of associated Gram matrices. *The minimum of the ranks of these Gram matrices turns out to be precisely the length of* $f$ (Theorem 2.4). This result provides the key to the general analysis of the length of polynomials undertaken in this paper.

In §3, we introduce the geometric-combinatorial method of "cages". This approach reveals some interesting connections between the theory of sums of squares in $A_n$ and the geometry of numbers. For fixed $n$, $m$ ($m$ even), a *cage* $C$ is, roughly speaking, a "convex" collection of $n$-ary $m$-ic monomials. To such a cage $C$, we can associate two families of forms:

(1.1) $F^+(C) := \{f : f \text{ is psd, containing only monomials in } C\}$,

(1.2) $F(C) := \{f \in F^+(C) : f \text{ is a sum of squares}\}$.

Using the second family, we can define the *pythagoras number* of the cage:

$$P(C) := \sup\{\text{length}(f) : f \in F(C)\}.$$

In case $C$ is the "full" cage $C_{n,m}$ consisting of all $n$-ary $m$-ic monomials, we recover, of course, the pythagoras number $P(n, m)$ introduced before.

For any cage $C$, there are three basic combinatorial invariants:

$l = l(C) :=$ number of monomials in $C$,

$e = e(C) :=$ number of even monomials in $C$,

$a = a(C) :=$ number of distinct "geometric means" of pairs of even monomials in $C$.

In §4 and onward, we use these invariants to get information on $F^+(C)$, $F(C)$, and $P(C)$. In §5, we characterize the "interiors" of $F^+(C)$, $F(C)$ (Proposition 5.4, 5.5), and show that these two spaces have topological dimensions, respectively, $l(C)$ and $a(C)$. In §4 and §6, we obtain upper and lower bounds for $P(C)$ in terms of $e = e(C)$ and $a = a(C)$:

Main Theorem (see (4.4) and (6.1)). *For any cage $C$, we have*

$$a/e \le \lambda \le P(C) \le \Lambda \le e,$$

*where* $\Lambda := (\sqrt{1 + 8a} - 1)/2$ *and* $\lambda := (2e + 1 - \sqrt{(2e+1)^2 - 8a})/2$.

In the case of the full cage $C = C_{n,m}$, one has $l(C) = \binom{n+m-1}{n-1}$, $e(C) = \binom{n+\frac{m}{2}-1}{n-1}$, and one can show that $a(C) = l(C)$ (Lemma 3.4). Thus, the theorem gives explicit upper and lower bounds for the pythagoras number $P(n, m)$.

If we fix $m$ and let $n$ vary, the theorem above implies that $P(n, m)$ has the order of magnitude of $n^{m/2}$ (see (6.4)). For instance, for quartic forms ($m = 4$), $P(n, 4)$ is roughly between $0.092n^2$ and $0.289n^2$ (see (6.5)), when $n$ is large. Of course, if we fix $n \ge 3$ and let $m$ vary, then $\sup\{P(n, m)\}$ is $\infty$ according to our earlier work **[CDLR]**.

The flexibility of using cages enables us to get sharper results for specific classes of forms. For instance, applying the Main Theorem above to *biquadratic* forms $\sum a_{ijkl} x_i x_j y_k y_l$ ($1 \le i, j \le n_1$, $1 \le k, l \le n_2$), we see that the pythagoras number for such forms is at most

$$(\sqrt{1 + 2n_1 n_2 (n_1 + 1)(n_2 + 1)} - 1)/2.$$

## §2. Sums of squares and Gram matrices

In this section, we show that the problem of expressing a form $f$ as a sum of squares of other forms and the problem of computing length$(f)$ can both be studied naturally from the viewpoint of linear algebra. The basic arguments needed from linear algebra will be collected here for later use.

We adopt the following multinomial notation: for $x = (x_1, \dots, x_n)$ and $\alpha = (\alpha_1, \dots, \alpha_n)$, write $x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$. A given $n$-ary $m$-ic can be written as $f(x) = \sum a_\alpha x^\alpha$, where $\alpha$ ranges over some set of $n$-tuples of nonnegative integers with $\alpha_1 + \cdots + \alpha_n = m$. Suppose $f$ is a sum of squares, say, $f = \sum_{i=1}^t h_i^2$. As we have indicated before, the $h_i$'s must be $m/2$-ics, say $h_i = \sum_\beta u_\beta^{(i)} x^\beta$ $(1 \le i \le t)$. Let $U_\beta = (u_\beta^{(1)}, \dots, u_\beta^{(t)}) \in \mathbf{R}^t$. Then

$$f = \sum_{i=1}^t h_i^2 = \sum_{i=1}^t \left(\sum_\beta u_\beta^{(i)} x^\beta\right)\left(\sum_{\beta'} u_{\beta'}^{(i)} x^{\beta'}\right) = \sum_\beta \sum_{\beta'} U_\beta \cdot U_{\beta'} x^{\beta+\beta'}. \tag{2.1}$$

[1] Hilbert's proof of this theorem expressed in the older language of algebraic geometry is understandably appreciated by few modern readers. For a proof of Hilbert's result written in the current mathematical terminology, see [Ra, Chapter 7], or the article of Swan [Sw] in these Proceedings.

Comparing coefficients, we get

$$a_\alpha = \sum_{\beta+\beta'=\alpha} U_\beta \cdot U_{\beta'}. \tag{2.2}$$

Conversely, if there exist $t$-vectors $\{U_\beta\} \subseteq \mathbf{R}^t$ which satisfy (2.2), then we can write $f$ as a sum of $t$ squares, using the coordinates of the $U_\beta$'s to construct the $h_i$'s. Thus, finding the length of $f$ becomes a question of finding a set of vectors $\{U_\beta\} \subseteq \mathbf{R}^t$ *for the smallest possible* $t$, such that the dot products $U_\beta \cdot U_{\beta'}$ satisfy certain linear equations, viz. (2.2).

Consider the dot product matrix $(U_\beta \cdot U_{\beta'})$ whose rows and columns are indexed by $\beta$, $\beta'$. We say that this is the *Gram matrix* associated with the sum of squares expression $f = \sum h_i^2$. In the literature, the characterization for such Gram matrices is well known. We shall recall this characterization in the form we need in later sections. For the sake of completeness, we shall also supply a proof.

**Proposition 2.3.** *Let $(v_{\beta\beta'})$ be a symmetric real matrix with associated quadratic form $Q(y) = \sum v_{\beta\beta'} y_\beta y_{\beta'}$. Then there exist real vectors $\{V_\beta\}$ with $V_\beta \cdot V_{\beta'} = v_{\beta\beta'}$ (for all $\beta$, $\beta'$) if and only if $Q(y)$ is psd. If this is the case, the smallest $t$ for which such vectors can be found in $\mathbf{R}^t$ is precisely* $\operatorname{rank}(v_{\beta\beta'})$ *(i.e., the rank of the quadratic form $Q$).*

**Proof.** If vectors $V_\beta = (v_\beta^{(1)}, \ldots, v_\beta^{(s)})$ exist with $V_\beta \cdot V_{\beta'} = v_{\beta\beta'}$, then $v_{\beta\beta'} = \sum_{i=1}^s v_\beta^{(i)} v_{\beta'}^{(i)}$. Thus,

$$Q(y) = \sum_{\beta,\beta'} \left( \sum_{i=1}^s v_\beta^{(i)} v_{\beta'}^{(i)} \right) y_\beta y_{\beta'} = \sum_{i=1}^s \left( \sum_\beta v_\beta^{(i)} y_\beta \right)^2 \geq 0,$$

so the rank of $Q$ is at most $s$. Conversely, suppose $Q$ is psd with rank $s$. Then

$$Q(y) = \sum v_{\beta\beta'} y_\beta y_{\beta'} = \sum_{i=1}^s \left( \sum_\beta v_\beta^{(i)} y_\beta \right)^2$$

for suitable real numbers $\{v_\beta^{(i)}\}$. Comparing the coefficients of $y_\beta y_{\beta'}$ on both sides, we get $v_{\beta\beta'} = \sum_{i=1}^s v_\beta^{(i)} v_{\beta'}^{(i)}$. Letting $V_\beta := (v_\beta^{(1)}, \ldots, v_\beta^{(s)})$, we have then $v_{\beta\beta'} = V_\beta \cdot V_{\beta'}$. QED

From our calculations at the beginning of this section and the above characterization of Gram matrices, we obtain immediately the following:

**Theorem 2.4.** (1) *Let $f(x) = \sum a_\alpha x^\alpha$, and $V = (v_{\beta\beta'})$ be a real symmetric matrix. The following statements are equivalent:*

*(A) $f$ is a sum of squares and $V$ is a Gram matrix associated to $f$ (with respect to some sums of squares expression $f = \sum h_i^2$);*



*(B) $V$ is psd and*

$$\sum_{\beta+\beta'=\alpha} v_{\beta\beta'} = a_\alpha \quad (\text{for all } \alpha). \tag{2.5}$$

(2) *If $f$ is a sum of squares, then*

$$\mathit{length}(f) = \min\{\operatorname{rank} V\} = \min\{\mathit{length}\ Q\}, \tag{2.6}$$

*where $V$ ranges over all Gram matrices associated to $f$, and $Q$ ranges over the quadratic forms associated to such $V$'s.*

Via this result, one is able to reduce the notion of length for *arbitrary* forms to the more familiar notion of the length of *quadratic* forms.

For later reference, we make the following observation on the Gram matrix of a nonnegative linear combination of forms (of the same degree).

**Lemma 2.7** (Semilinearity). *Suppose $f = \sum h_i^2$ has associated Gram matrix $U$, and $g = \sum k_j^2$ has associated Gram matrix $V$. Then for any $c, d \geq 0$, the following sum of squares expression*

$$cf + dg = \sum (\sqrt{c} h_i)^2 + \sum (\sqrt{d} k_j)^2 \tag{2.8}$$

*has associated Gram matrix $cU + dV$.*

**Proof.** Suppose $\{U_\beta\}$, $\{V_\beta\}$ are the (row) vectors arising from the $h_i$'s and $k_j$'s respectively. Then $\{(\sqrt{c}U_\beta, \sqrt{d}V_\beta)\}$ are the (row) vectors arising from $\{\sqrt{c}h_i, \sqrt{d}k_j\}$. Since

$$(\sqrt{c}U_\beta, \sqrt{d}V_\beta) \cdot (\sqrt{c}U_{\beta'}, \sqrt{d}V_{\beta'}) = cU_\beta \cdot U_{\beta'} + dV_\beta \cdot V_{\beta'},$$

the Gram matrix associated with the expression (2.8) is clearly $cU+dV$. QED

Let $f = \sum_{i=1}^t h_i^2$ be a given expression of $f$ as a sum of $t$ squares, and, as before, write $h_i = \sum_\beta u_\beta^{(i)} x^\beta$ and $U_\beta = (u_\beta^{(1)}, \ldots, u_\beta^{(t)})$. We can always derive other expressions of $f$ as a sum of $t$ squares by the following procedure. Let $(c_{ij})$ be a $t \times t$ real orthogonal matrix, and let $\{\tilde{h}_i = \sum_\beta \tilde{u}_\beta^{(i)} x^\beta\}$ by defined by

$$(\tilde{h}_1, \ldots, \tilde{h}_t) = (h_1, \ldots, h_t)(c_{ij}).$$

Then $\sum_{i=1}^t \tilde{h}_i^2 = \sum_{i=1}^t h_i^2 = f$, and

$$\tilde{h}_j = \sum_i h_i c_{ij} = \sum_\beta \left( \sum_i u_\beta^{(i)} c_{ij} \right) x^\beta$$

implies that $\tilde{u}_\beta^{(j)} = \sum_i u_\beta^{(i)} c_{ij}$. Thus, we may think of the new $\widetilde{U}_\beta$'s as the old vectors $U_\beta$'s expressed in coordinates with respect to a new orthonormal basis of $\mathbf{R}^t$. In particular, $\widetilde{U}_\beta \cdot \widetilde{U}_{\beta'} = U_\beta \cdot U_{\beta'}$, and so the two expressions

$$f = \sum_{i=1}^t h_i^2, \qquad f = \sum_{i=1}^t \tilde{h}_i^2 \tag{2.9}$$

of $f$ as sums of $t$ squares have the same Gram matrix. In the sequel, we shall say that the second expression for $f$ above is obtained from the first *by an orthogonal transformation* or that the two expressions are *orthogonally equivalent*.

PROPOSITION 2.10. *A sum of squares expression* $f = \sum_{i=1}^t g_i^2$ *can be obtained from* (2.9) *by an orthogonal transformation if and only if the two expressions have the same Gram matrix.*

PROOF. We need only prove the "if" part. Write $g_i = \sum_\beta v_\beta^{(i)} x^\beta$, $V_\beta = (v_\beta^{(1)}, \dots, v_\beta^{(t)})$, and assume that

$$U_\beta \cdot U_{\beta'} = V_\beta \cdot V_{\beta'} \quad \text{for all } \beta, \beta'. \tag{2.11}$$

Let $A$ (resp., $B$) be the subspace of $\mathbf{R}^t$ spanned by the $U_\beta$'s (resp., $V_\beta$'s). From (2.11), it is easy to show that the map given by

$$\sum \varepsilon_\beta U_\beta \in A \mapsto \sum \varepsilon_\beta V_\beta \in B$$

is well defined, and is an isometry from $A$ to $B$. By Witt's Extension Theorem **[La**, p. 26] (which is obvious in the present context), this extends to an isometry of $\mathbf{R}^t$. Therefore, we may view the $V_\beta$'s as the coordinate vectors of the $U_\beta$'s with respect to some orthonormal basis of $\mathbf{R}^t$. Reversing the calculation given before the statement of (2.10), we see that the expression $f = \sum_{i=1}^t g_i^2$ is obtainable by an orthogonal transformation of the expression $f = \sum_{i=1}^t h_i^2$. QED

In view of the Proposition above, *the classification of the Gram matrices of* $f$ *corresponds to the classification of the expression of* $f$ *as sums of squares up to orthogonal equivalence.*

For ease of notation, we shall often label the $\beta$'s as $\beta_1, \beta_2, \dots, \beta_e$, and write accordingly $U_i$ for $U_{\beta_i}$, etc. For any such fixed labelling, if $f$ is a sum of squares, we can always write it as a sum of squares in a specific way, as in the following result.

COROLLARY 2.12. *Let* $f = \sum_{i=1}^t h_i^2$. *Then, after an orthogonal transformation, we can write* $f = \sum_{i=1}^t g_i^2$ *such that* $x^{\beta_1}$ *occurs only in* $g_1$, $x^{\beta_2}$ *occurs only in* $g_1$, $g_2$, *etc.*

PROOF. By the Gram-Schmidt Orthonormalization Process, it is possible to choose an orthonormal basis $\{e_1', \dots, e_t'\}$ for $\mathbf{R}^t$ such that, for each $i$, the vector $U_i$ lies in the span of $\{e_1', \dots, e_i'\}$. By the discussion preceding (2.10), the new orthonormal basis $\{e_1', \dots, e_t'\}$ determines an orthogonal transformation of $f = \sum_{i=1}^t h_i^2$ into a new expression $f = \sum_{i=1}^t g_i^2$. The new vectors $V_1, V_2, \dots$ associated with this expression now have the form $(*, 0, 0, \dots, 0)$, $(*, *, 0, \dots, 0)$, $\dots$, so for any $i$, $x^{\beta_i}$ can only occur in $g_1, g_2, \dots, g_i$, as desired. QED



This result will be very useful later in obtaining lower bounds for the length of sums of squares of forms; see (6.1).

We shall now address the following important question: *Suppose a form* $f$ *is known to be a sum of squares. What kind of algebraic object is the set of Gram matrices associated to* $f$? Let $f = \sum_{i=1}^t h_i^2$ be a specific expression of $f$ as a sum of $t$ squares, with an associated Gram matrix $(u_{\beta\beta'})$. Then any solution for (2.5) is given by $v_{\beta\beta'} = u_{\beta\beta'} + \varepsilon_{\beta\beta'}$ where $(\varepsilon_{\beta\beta'})$ satisfy the *homogeneous* linear equations $\sum_{\beta+\beta'=\alpha} \varepsilon_{\beta\beta'} = 0$. Of course, we also want $\varepsilon_{\beta\beta'} = \varepsilon_{\beta'\beta}$ to insure the symmetry of $(v_{\beta\beta'})$. Note that the above linear equations are linearly independent. Thus, if the indexing set for $\alpha$ (resp., $\beta$) has $a$ (resp., $e$) elements, then $(v_{\beta\beta'})$ can be expressed linearly by $\delta := e(e+1)/2 - a$ independent parameters $\lambda_1, \dots, \lambda_\delta$. In the corresponding parameter space $\mathbf{R}^\delta$, *the Gram matrices associated with* $f$ *is then a closed semialgebraic set, defined by polynomial inequalities in* $\lambda_1, \dots, \lambda_\delta$ *dictated by the condition that the principal minors in the Gram matrix be all nonnegative.* These inequalities could considerably cut down the "degree of freedom" for the Gram matrices. For instance, it may happen that $\delta$ is a large number, but $f$ has a *unique* Gram matrix $V$. This would, indeed, be a very fortunate case, because then length($f$) is immediately computed by determining rank $V$, and we further know that $f$ has a *unique* expression as a sum of squares, up to orthogonal transformation.

EXAMPLE 2.13. Let $f(x, y) \in \mathbf{R}[x, y]$ be a product of $m$ distinct irreducible binary quadratic forms. Classify the different expressions of $f$ as a sum of two squares, up to orthogonal equivalence.

Without loss of generality, we may assume that the coefficient of $x^{2m}$ in $f$ is 1. We shall first construct $2^{m-1}$ essentially different ways of writing $f$ as a sum of two squares, and then show that these are *all* the possible ways, up to orthogonal equivalence.

Write $f(x, y) = \prod_{j=1}^m (x - \alpha_j y)(x - \overline{\alpha}_j y)$ in $\mathbf{C}[x, y]$, and define real forms $P(x, y)$, $Q(x, y)$ by $P + iQ = \prod_{j=1}^m (x - \alpha_j y)$. In this way, we obtain *one* way of writing $f$ as a sum of two squares, namely, $f = P^2 + Q^2$. If we replace one or more of the $\alpha_j$'s by $\overline{\alpha}_j$, we obtain another pair of real forms, say $P'$, $Q'$, such that $f = P'^2 + Q'^2$. Ostensibly, this gives $2^m$ decompositions. Up to orthogonal equivalence, however, it is easy to see that we get only $2^{m-1}$ decompositions. (Since the coefficients of $x^m$ in $P$ and $P'$ are both 1, the two expressions $f = P^2 + Q^2$ and $f = P'^2 + Q'^2$ are orthogonally equivalent iff $Q' = \pm Q$.)

It is now easy to see that we have exhausted all possible ways of writing $f$ as a sum of two squares, up to orthogonal equivalence. For, let $f = \tilde{P}^2 + \tilde{Q}^2 = (\tilde{P} + i\tilde{Q})(\tilde{P} - i\tilde{Q})$. By the theorem of unique factorization, we see that $\tilde{P} + i\tilde{Q} = \lambda(P' + iQ')$ for some $\lambda \in \mathbf{C}$, where $P'$, $Q'$ are as above. Clearly, $\lambda\bar{\lambda} = 1$, so writing $\lambda = a + bi$, we have $a^2 + b^2 = 1$, i.e., $\begin{pmatrix} a & b \\ -b & a \end{pmatrix}$

is an orthogonal matrix. Since $(\widetilde{P}, \widetilde{Q}) = (P', Q')\left(\begin{smallmatrix} a & b \\ -b & a\end{smallmatrix}\right)$, the expression $f = \widetilde{P}^2 + \widetilde{Q}^2$ is orthogonally equivalent to the expression $f = P'^2 + Q'^2$ constructed in the last paragraph. In view of Proposition 2.10, the analysis above shows that $f$ *has exactly* $2^{m-1}$ *Gram matrices*; each of these Gram matrices has rank 2 since $f$ is not a perfect square.

For a concrete example, let $m = 2$ and $f = (x^2+y^2)(x^2+4y^2)$. A general Gram matrix for $f$ is easily seen to be of the form

$$G_r = \begin{pmatrix} 1 & 0 & r \\ 0 & 5-2r & 0 \\ r & 0 & 4 \end{pmatrix} \qquad (|r| \leq 2),$$

where the rows (and columns) correspond to the monomials $x^2$, $xy$, and $y^2$, in that order. Since $\det(G_r) = (r+2)(r-2)(2r-5)$, the two Gram matrices of rank 2 are given by the parameter values $r = \pm 2$; namely,

$$\begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \\ 2 & 0 & 4 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & 0 & -2 \\ 0 & 9 & 0 \\ -2 & 0 & 4 \end{pmatrix}.$$

The corresponding expressions of $f$ as sums of two squares are

$$f = (x^2+2y^2)^2 + (xy)^2 \quad \text{and} \quad f = (x^2-2y^2)^2 + (3xy)^2,$$

as predicted by the proof of (2.13). (Incidentally, these are precisely the expressions obtained by applying the 2-square identity to $(x^2+y^2)(x^2+4y^2)$ in the two obvious ways.)

While we are on the topic of psd binary forms, the following remarkable application of the Gram matrix method seems worth mentioning:

**Proposition 2.14.** *Let* $(a_{ij})_{0\leq i,j\leq n}$ *be a psd real symmetric matrix. Then there exists a psd real symmetric matrix* $(b_{ij})$ *of rank* $\leq 2$ *such that* $\sum_{i+j=r} a_{ij} = \sum_{i+j=r} b_{ij}$ *for every* $r$.

**Proof.** Specialize the psd quadratic form $\sum a_{ij}x_ix_j$ by $x_i \mapsto t^i$, where $t$ is a real variable. We get a psd polynomial $f(t) = \sum a_{ij}t^{i+j} = \sum_r(\sum_{i+j=r} a_{ij})t^r$ which can then be written as $p(t)^2 + q(t)^2$. The Gram matrix $(b_{ij})$ associated with this expression of $f$ as a sum of two squares is the matrix we want, since $\sum_{i+j=r} b_{ij} = \sum_{i+j=r} a_{ij}$ follows from (2.2). (Of course, we are applying the Gram matrix method here in the *inhomogeneous* case.) QED

## §3. The method of cages

To study $n$-ary $m$-ics more systematically, we set up some notations for the geometry of $\mathbf{R}^n$. Throughout the rest of the paper, we write

$$C_{n,m} = \left\{(c_1, \ldots, c_n) : 0 \leq c_i \in \mathbf{R}, \sum c_i = m\right\}.$$



This is an $(n-1)$-simplex with vertices $(m, 0, \ldots, 0)$, $(0, m, \ldots, 0)$, $\ldots$, and $(0, 0, \ldots, m)$. An $n$-tuple $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbf{R}^n$ is called a *lattice point* if each $\alpha_i \in \mathbf{Z}$; $\alpha$ is called an *even lattice point* if each $\alpha_i \in 2\mathbf{Z}$. (Lattice points will be consistently denoted by Greek letters.) For any set $C \subseteq \mathbf{R}^n$, we shall write

$$L(C) := C \cap \mathbf{Z}^n \text{ (the set of lattice points in } C\text{)},$$
$$E(C) := C \cap (2\mathbf{Z})^n \text{ (the set of even lattice points in } C\text{)}.$$

An $n$-ary $m$-ic $f$ is written as $f = \sum a_\alpha x^\alpha$, where $\alpha$ ranges over (some subset of) $L(C_{n,m})$. Following **[Re₁]**, we define the *cage* of $f$, denoted $C(f)$, to be the (closed) convex hull of those $\alpha$ for which $a_\alpha \neq 0$ ($C(f)$ is also known as the "Newton polytope" of $f$). We recall the following basic fact obtained earlier by the third author:

**Theorem 3.1** [**Re₁**, Lemma in §3]. *If $f$ above is psd, then the cage $C(f)$ is the convex hull of a finite set of even lattice points (i.e., $C(f)$ is the convex hull of $E(C(f))$).*

Conversely, if $C$ is the convex hull of some nonempty set of even lattice points in $C_{n,m}$ ($m$ necessarily even), then $C$ is indeed the cage of a suitable psd $n$-ary $m$-ic. For instance, $C$ is the cage of the sum of squares $f = \sum x^\alpha$, where $\alpha$ ranges over $E(C)$. In the following, we shall call any such $C$ a cage in $C_{n,m}$, without reference to any specific psd form (or forms) $f$. Observe that, if $f$ is psd, then $C(f) = C_{n,m}$ iff $x_1^m, \ldots, x_n^m$ all have positive coefficients in $f$; that is, iff none of the standard unit vectors is a zero of $f$.

In this section, we shall begin the development of a general theory for sums of squares with respect to *any* cage $C$ in $C_{n,m}$ ($m$ even). Of course, $C = C_{n,m}$ will be a special case. The point is that the theory developed for a general cage $C$ will have greater flexibility in applications, and will in fact yield stronger results in most cases.

For $C$ any cage in $C_{n,m}$, we adopt the following notations and terminology. The set $L(C)$ of lattice points in $C$ is called the *frame* of the case $C$ (see **[Re₁]**). The set $\frac{1}{2}C = \{\frac{1}{2}c : c \in C\}$ is called the *half-cage* of $C$. Note that from the definitions, we have $L(\frac{1}{2}C) = \frac{1}{2}E(C)$. For any $\alpha \in L(C)$, let $D(\alpha) = D_C(\alpha)$ denote the set of ordered pairs $(\beta, \beta')$ of lattice points in the half-cage $\frac{1}{2}C$ such that $\beta + \beta' = \alpha$ (vector addition). Then we define:

$$A(C) = \{\alpha \in L(C) : D(\alpha) \neq \varnothing\}. \tag{3.2}$$

This consists of lattice points $\alpha \in C$ which decompose in at least one way as a sum of two lattice points in the half-cage. Since $L(\frac{1}{2}C) = \frac{1}{2}E(C)$, we see that $A(C)$ is just the set of "averages" (hence the notation "$A$") of pairs of even lattice points in $C$. Note that we have the inclusions $L(C) \supseteq A(C) \supseteq E(C)$ (the latter is seen by viewing $\beta \in E(C)$ as $\beta = \frac{1}{2}(\beta + \beta)$). These inclusions may fail to be equalities in general. However, we shall show below (cf. (3.4)) that $L(C) = A(C)$ for the "full" cage $C = C_{n,m}$. We have also shown that $L(C) = A(C)$ for *any* ternary cage $C \subseteq C_{3,m}$, though this result will not be presented here.

**Permanent Notations (3.3).** For any cage $C \subseteq C_{n,m}$, we shall write $F^+(C)$ for the family of $n$-ary $m$-ics $f$ such that $f$ is psd and $C(f) \subseteq C$. We write $F(C)$ for the subfamily $\{f \in F^+(C): f \text{ is a sum of squares}\}$. The form $f_C := \sum_{\alpha\in E(C)} x^\alpha \in F(C)$ will be called the *principal form* of the cage. The basic numerical invariants for $C$ are: $l = l(C)$, $a = a(C)$ and $e = e(C)$, which are, respectively, the cardinalities of $L = L(C)$, $A = A(C)$, and $E = E(C)$. Obviously, we have the inequalities $e \le a \le e(e+1)/2$. The number $\delta = \delta(C) := e(e+1)/2 - a \ge 0$ will be called the *defect* of the cage $C$; the significance of $\delta$ has already been hinted at in the discussion following Theorem 2.4. For the full cage $C = C_{n,m}$, we write

$$L_{n,m} = L(C), \quad A_{n,m} = A(C), \quad \text{and} \quad E_{n,m} = E(C).$$

Then $l(n, m) = l(C)$ is the number of $m$-ic monomials in $x_1, \dots, x_n$, i.e., $l(n, m) = \binom{n+m-1}{n-1}$. In the same vein, $e(n, m) = e(C)$ is the number of $m/2$-ic monomials, i.e., $e(n, m) = \binom{n+\frac{1}{2}m-1}{n-1}$. The following lemma computes the remaining number: $a(n, m) = \binom{n+m-1}{n-1}$.

LEMMA 3.4. *For $C = C_{n,m}$ ($m$ even), we have $L_{n,m} = A_{n,m}$.*

PROOF. We must show that any $n$-tuple $\alpha = (\alpha_1, \dots, \alpha_n) \in L_{n,m}$ can be written as $\alpha = \frac{1}{2}(\gamma + \gamma')$, where $\gamma, \gamma' \in E_{n,m}$. Since $m$ is even, we must have an even number of odd components among $\{\alpha_i\}$. Now define $\gamma, \gamma'$ as follows:

$$\begin{cases} \gamma_i = \gamma_i' = \alpha_i & \text{if } \alpha_i \text{ is even},\\ \gamma_i = \alpha_i + (-1)^k,\ \gamma_i' = \alpha_i - (-1)^k & \text{if } \alpha_i \text{ is the } k\text{th odd component.}\end{cases}$$

Since $\alpha_i$ odd implies that $\alpha_i \ge 1$, we have $\gamma_i, \gamma_i' \ge 0$. Finally,

$$\sum \gamma_i = \sum \alpha_i + \sum_k (-1)^k = m,$$

since there are an even number of odd $\alpha_i$'s. Similarly $\sum \gamma_i' = m$, so $\gamma, \gamma' \in E_{n,m}$. We have now the desired decomposition $\alpha = \frac{1}{2}(\gamma + \gamma')$. QED

Going back to an arbitrary cage $C \subseteq C_{n,m}$, we shall now supply the motivation for the definitions of $D(\alpha)$ and $A(C)$ given earlier. Let $f = \sum_{\alpha\in L(C)} a_\alpha x^\alpha \in F(C)$, say $f = \sum_{i=1}^t h_i^2$ ($h_i = m/2$-ics). We recall the following basic relationship between $C(f)$ and $C(h_i)$:

THEOREM 3.5 **Re**$_1$, Theorem 1]. *For any $i$, $C(h_i) \subseteq \frac{1}{2}C(f)$.*



According to this result, we can write $h_i = \sum_{\beta\in L(\frac{1}{2}C)} u_\beta^{(i)} x^\beta$. As in §2, write $U_\beta = (u_\beta^{(i)}, \dots, u_\beta^{(t)})$. By our earlier calculation, we have

$$a_\alpha = \sum_{(\beta,\beta')\in D(\alpha)} U_\beta \cdot U_{\beta'} \tag{3.6}$$

for any $\alpha \in L(C)$. This basic equation is the principal motivation for introducing the sets $D(\alpha)$. *Note that this sum is nonempty iff $\alpha \in A(C)$.*

As in §2, we say that $U = (U_\beta \cdot U_{\beta'})$ is the *Gram matrix* associated to the sum of squares expression $f = \sum h_i^2$. Strictly speaking, this Gram matrix depends on the choice of $C$ (as well as on the equation $f = \sum h_i^2$). However, if $C'$ is another cage such that $C \subseteq C'$, we see easily that the new Gram matrix of $f = \sum h_i^2$ with respect to $C'$ is just $\begin{pmatrix} U & 0 \\ 0 & 0\end{pmatrix}$ (since the "new vectors" $U_\gamma$ with $\gamma \in L(\frac{1}{2}C')\backslash L(\frac{1}{2}C)$ are all zero). Thus, for all intents and purposes, we may regard the Gram matrix of $f = \sum h_i^2$ as "independent" of the choice of $C$. As an example, consider the principal form of the cage $f_C := \sum_{\alpha\in E(C)} x^\alpha$. With respect to *this* sum of squares expression, one checks easily that *the associated Gram matrix is the identity matrix* (*of size* $e(C) \times e(C)$).

Going back to the equation (3.6), we can now record some nontrivial *necessary* conditions for a form $f$ to be a sum of squares.

PROPOSITION 3.7. *Let $f = \sum_\alpha a_\alpha x^\alpha \in F(C)$; say $f = \sum_i h_i^2$. Then*
(1) *For any $\alpha \in L(C)\backslash A(C)$, we have $a_\alpha = 0$.*
(2) *For any $\alpha \in E(C)$ such that $|D(\alpha)| = 1$, we have $a_\alpha \ge 0$. If in fact $a_\alpha = 0$, then the monomial $x^{\alpha/2}$ cannot occur in any of the $h_i$'s.*

PROOF. (1) If $\alpha \notin A(C)$, then (3.6) is an empty sum, so $a_\alpha = 0$.
(2) By hypothesis, $D(\alpha) = \{(\beta, \beta)\}$, where $\beta = \alpha/2 \in L(\frac{1}{2}C)$. Thus, $a_\alpha = U_\beta \cdot U_\beta \ge 0$. If $a_\alpha = 0$, then we must have $U_\beta = 0$; this means that $x^\beta$ cannot occur in any of the $h_i$. QED

The simplest example of an even lattice point $\alpha \in E(C)$ with $|D(\alpha)| = 1$ is when $\alpha = (m, 0, \dots, 0)$ (assuming it lies in $C$). For this choice of $\alpha$, (2) above says that $x_1^m$ occurs with non-negative coefficient in $f$, and that if $x_1^m$ does not occur in $f$, then $x_1^{m/2}$ cannot occur in any of the $h_i$'s. This is, of course, obvious upon evaluation of $f = \sum h_i^2$ at $(1, 0, \dots, 0)$.

The above type of necessary conditions derived for sums of squares may be viewed as the general abstract formulation of the "term inspection" method used by the first two authors in **[CL$_1$]**. For illustration, let us reiterate two of the key examples in **[CL$_1$]**, using the current notations.

EXAMPLE 3.8. For the psd quaternary quartic

$$f(w, x, y, z) = w^4 + x^2y^2 + y^2z^2 + x^2z^2 - 4wxyz,$$

let $C = C(f)$, which is the tetrahedron with vertices $(4, 0, 0, 0)$, $(0, 2, 2, 0)$, $(0, 0, 2, 2)$, $(0, 2, 0, 2)$, and $E(C)$ consists precisely of these four points. By checking out the averages of pairs of these four points, we see that $|A(C)| = 10$, and that $(1, 1, 1, 1) \notin A(C)$. But $(1, 1, 1, 1) \in L(C)$; in fact, $L(C)$ consists of $A(C)$ plus the one point $(1, 1, 1, 1)$, so in this example, $e = 4$, $a = 10$, and $l = 11$. By (3.7)(1), the occurrence of the term $wxyz$ (corresponding to $(1, 1, 1, 1)$) shows that $f$ cannot be a sum of squares.

EXAMPLE 3.9. For the psd ternary sextic

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

again let $C = C(f)$. The sets $L(C)$, $L(\frac{1}{2}C)$ and $A(C)$ are easily found by inspection, as follows:

$$(3.10)\quad \begin{aligned} E(C) &= \{(4, 2, 0), (0, 4, 2), (2, 0, 4), (2, 2, 2)\} \\ L(\tfrac{1}{2}C) &= \tfrac{1}{2}E(C) = \{(2, 1, 0), (0, 2, 1), (1, 0, 2), (1, 1, 1)\} \\ L(C) &= A(C) = \\ &E(C) \cup \{(1, 2, 3), (2, 3, 1), (3, 1, 2), (1, 3, 2), (2, 1, 3), (3, 2, 1)\}. \end{aligned}$$

In particular, $D(2, 2, 2) = \{(\beta, \beta)\}$, where $\beta = (1, 1, 1)$. But $a_{(2,2,2)} = -3$, so (3.7)(2) shows that $f$ cannot be a sum of squares. In this example, $e = 4$, $a = l = 10$.[2] We can illustrate the sets in (3.10) by a 2-dimensional picture (the projection of $C$ onto the first two coordinates[3] (see Figure 1).

EXAMPLE 3.11. To show how the Gram matrix method works, let us compute, for example, the length of $f(v, w, x, y, z) = (v^2 + w^2)(x^2 + y^2 + z^2)$ and some different ways of expressing $f$ as a sum of squares. The 2-square identity implies that length$(f) \le 4$, but we shall show length$(f) = 4$ independently. Let $C = C(f)$. Then $E(C) = \{(\gamma, \gamma')\}$, where $\gamma = (2, 0)$ or $(0, 2)$, and $\gamma' = (2, 0, 0)$, $(0, 2, 0)$ or $(0, 0, 2)$. Using the quadratic monomials $vx, wy, wz, vy, vz, wx$ (in that order) to index rows and columns, an arbitrary Gram matrix for $f$ is easily seen to be of the form:

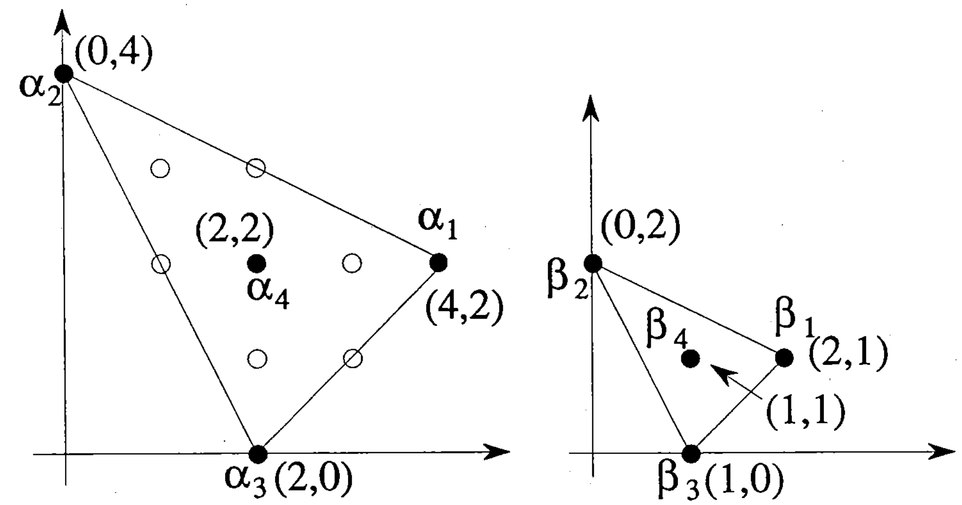
$$(3.12)\qquad \begin{pmatrix} 1 & r & s & 0 & 0 & 0 \\ r & 1 & 0 & 0 & -t & 0 \\ s & 0 & 1 & t & 0 & 0 \\ 0 & 0 & t & 1 & 0 & -r \\ 0 & -t & 0 & 0 & 1 & -s \\ 0 & 0 & 0 & -r & -s & 1 \end{pmatrix}.$$

Assume this matrix has rank 3; then any $4\times4$ minor has zero determinant. Applying this to the minor formed by the first, second, fourth and sixth rows and columns, we see that $r = \pm1$, and similarly, $s = \pm1$, $t = \pm1$. But

[2]As we have pointed out in the paragraph preceding (3.3), $a = l$ holds for *any* cage in the ternary case.

[3]This is a useful technique. In general, since $C \subseteq C_{n,m}$, the last coordinate is redundant, so it suffices to look at the "picture" obtained by projecting $C$ onto the first $n - 1$ coordinates.






The Cage $C = C(f)$, $E(C) = \{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}$ The half-cage $\frac{1}{2}C$, $L(\frac{1}{2}C) = \{\beta_1, \beta_2, \beta_3, \beta_4\}$

FIGURE 1

then the minor formed by the first four rows and columns has determinant $-1$, a contradiction. Thus, length$(f) \ge 4$. Setting $t = 0$, $r = \cos\theta$, and $s = \sin\theta$, (3.12) decomposes into a direct sum of two psd $3 \times 3$ matrices of rank 2, so we get an infinite family of Gram matrices for $f$ of rank 4. An easy realization of these Gram matrices by dot products in $\mathbf{R}^4$ is given by

$$\begin{cases} U_{vx} = (1, 0, 0, 0), \\ U_{wy} = (\cos\theta, \sin\theta, 0, 0), \\ U_{wz} = (\sin\theta, -\cos\theta, 0, 0), \\ U_{vy} = (0, 0, \cos\theta, \sin\theta), \\ U_{vz} = (0, 0, \sin\theta, -\cos\theta), \\ U_{wx} = (0, 0, -1, 0). \end{cases}$$

This leads to a one-parameter family of expressions of $f$ as a sum of four squares

$$(3.13)\qquad \begin{aligned} f = {} & [vx + (\cos\theta)wy + (\sin\theta)wz]^2 + [(\sin\theta)wy - (\cos\theta)wz]^2 \\ & + [(\cos\theta)vy + (\sin\theta)vz - wx]^2 + [(\sin\theta)vy - (\cos\theta)vz]^2. \end{aligned}$$

Note that for $0 \le \theta < 2\pi$, the family above gives *orthogonally inequivalent representations* of $f$ as a sum of four squares, since different $\theta$'s in the range $[0, 2\pi)$ give rise to different Gram matrices. Concerning this point, let us make two further observations:

**3.14.** Our definition of orthogonal equivalence of representations of a form as a sum of squares is not to be confused with the orthogonal equivalence of *bilinear pairings* defined by Yuzvinsky **[Yu]** and Bier-Schwardmann **[BS]**. In fact, though the representations of $f$ in (3.13) as sums of four squares are inequivalent in our sense, the corresponding bilinear maps $B_\theta : \mathbf{R}^2 \times \mathbf{R}^3 \to \mathbf{R}^4$ given by the four expressions in brackets in (3.13) are mutually equivalent in the sense of Yuzvinsky and Bier-Schwardmann. (This can be easily seen from the fact that we can also derive (3.13) from the 2-square identity, together with an orthogonal transformation on $y^2 + z^2$.)

**3.15.** The example given in (3.11) shows that a form of length 4 may have infinitely many Gram matrices of rank 4. On the other hand, the ideas used in Example (2.13) show that a form of length 2 can have at most a finite number of Gram matrices of rank 2. It is, therefore, natural to ask: *if a form* $g$ *has length* 3, *must* $g$ *have only finitely many Gram matrices of rank* 3? While this turned out to be the case in the few examples we have chosen to compute, a general answer to the question has so far eluded us.

## §4. Upper bounds for the pythagoras number

To any case $C \subseteq C_{n,m}$, we shall associate the following positive integer, called the *pythagoras number* of the cage:

$$P(C) := \sup\{\text{length}(f) : f \in F(C)\}, \tag{4.1}$$

where $F(C)$ is as defined in (1.2). Clearly, if $C'$ is another cage such that $C' \subseteq C$, then $P(C') \le P(C)$. We observe that the pythagoras number of a cage is always finite; in fact we have the following bound.

Proposition 4.2. *For any cage* $C \subseteq C_{n,m}$, $P(C) \le e(C)$.

Proof. Let $f = \sum h_i^2 \in F(C)$, with associated Gram matrix indexed by $\beta, \beta' \in \frac{1}{2}E(C)$. This square matrix has size $|E(C)| = e(C)$, so its rank is $\le e(C)$. By (2.6), we have $\text{length}(f) \le e(C)$. QED

Following the practice in (3.3), we shall write $P(n, m) := P(C_{n,m})$. By classical results on binary forms and quadratic forms, we have $P(2, m) = 2$, and $P(n, 2) = n$. Hilbert's famous result on ternary forms **[Hi]** gives $P(3, 4) = 3$. From (4.2), we have the upper bounds $P(n, m) \le e(n, m) \le l(n, m)$. The weaker bound $P(n, m) \le l(n, m) = \binom{n+m-1}{n-1}$ was implicit in **[Hi]** and explicitly noted by R. M. Robinson in **[Ro]**; the sharper bound $P(n, m) \le e(n, m) = \binom{n+\frac{m}{2}-1}{n-1}$ was first announced in **[CL$_2$]**. Our next goal is to show that, even for a general cage $C \subseteq C_{n,m}$, the upper bound $P(C) \le e(C)$ for the pythagoras number in (4.2) can be considerably improved. We first establish an elementary lemma on quadratic forms:

Lemma 4.3. *Let* $q(z_1, \dots, z_t)$ *be a nonzero real quadratic form. Then there exists* $\lambda \in \mathbf{R}$ *such that* $z_1^2 + \cdots + z_t^2 - \lambda q(z_1, \dots, z_t)$ *is psd with rank* $\le t - 1$.

PROOF. On the sphere $z_1^2+\cdots+z_t^2=1$, let $M=\max|q(z_1,\ldots,z_t)|>0$. Then for any $(a_1,\ldots,a_t)\in\mathbf{R}^t$,

$$-M(a_1^2+\cdots+a_t^2)\le q(a_1,\ldots,a_t)\le M(a_1^2+\cdots+a_t^2),$$

and hence the quadratic forms

$$q'(z_1,\ldots,z_t):=z_1^2+\cdots+z_t^2\pm M^{-1}q(z_1,\ldots,z_t)$$

are both psd. Choosing $\lambda=\pm M^{-1}$ with the sign depending on whether $q$ assumes the value $M$ or $-M$ on the unit sphere, we can ensure that $q'=(z_1^2+\cdots+z_t^2)-\lambda q$ has a nontrivial zero, so that the rank of $q'$ is $\le t-1$. QED

THEOREM 4.4. *For any cage $C\subseteq C_{n,m}$, let $P=P(C)$, $a=a(C)$, and $e=e(C)$. Then $P(P+1)/2\le a$. Let $\Lambda=\Lambda(C):=(\sqrt{1+8a}-1)/2$ be the positive root of the quadratic equation $G(x):=x^2+x-2a=0$. Then $P\le\Lambda\le e$ (improving* (4.2)*).*

PROOF. Suppose we have shown $P(P+1)/2\le a$. Then clearly $P$ cannot exceed the positive root $\Lambda$ of the quadratic equation $G(x)=0$. Also, $G(e)=e(e+1)-2a\ge 0$ by our earlier observation (see (3.3)), so we must have $\Lambda\le e$.

We shall now prove the inequality $P(P+1)/2\le a$. Let $f=\sum_{i=1}^t h_i^2\in F(C)$, with associated Gram matrix $(u_{\beta\beta'})$. By (4.2), we may assume that this sum of squares expression has been chosen so that $t\le e$. If $t(t+1)/2>a$, *we shall show that* length$(f)\le t-1$. Let $Q_u(y)=\sum u_{\beta\beta'}y_\beta y_{\beta'}$. We may assume that rank $Q_u=t$ for otherwise we would already have length$(f)\le\operatorname{rank}(Q_u)<t$ (by (2.4)). Let $v_{\beta\beta'}=u_{\beta\beta'}-\lambda\varepsilon_{\beta\beta'}$ and $Q_v(y)=Q_u(y)-\lambda Q_\varepsilon(y)$, where $\lambda\in\mathbf{R}$ is to be specified, and $\{\varepsilon_{\beta\beta'}\}$ are subject to the conditions

$$\varepsilon_{\beta\beta'}=\varepsilon_{\beta'\beta}\quad\text{and}\quad\sum_{\beta+\beta'=\alpha}\varepsilon_{\beta\beta'}=0\qquad(\forall\alpha\in A(C)).$$

This amounts to a linear system with $a$ equations and $e(e+1)/2$ unknowns. Since $Q_u(y)$ is psd of rank $t$, there is an invertible linear change of variables $y=Mz$ such that $Q_u(y)=z_{\beta_1}^2+\cdots+z_{\beta_t}^2$. Since $Q_v(y)=Q_u(y)-\lambda Q_\varepsilon(y)$, we now impose the restriction that $Q_\varepsilon(y)$ can only involve the variables $z_{\beta_1},\ldots,z_{\beta_t}$. The general coefficients of $z_\beta z_{\beta'}$ in $Q_\varepsilon(y)$ is a linear combination of the $\varepsilon_{\beta\beta'}$'s, so if we set $\frac12 e(e+1)-\frac12 t(t+1)$ of these equal to zero, we have now imposed, altogether, $\frac12 e(e+1)-\frac12 t(t+1)+a$ homogeneous linear conditions on $\frac12 e(e+1)$ unknowns. Thus, if $\frac12 t(t+1)>a$, there is a nontrivial solution $\{\varepsilon_{\beta\beta'}\}$. Now let $q(z_{\beta_1},\ldots,z_{\beta_t})=Q_\varepsilon(Mz)\ne 0$ and apply Lemma 4.3. This enables us to find $\lambda\in\mathbf{R}$ such that $Q_v(y)=Q_u(y)-\lambda Q_\varepsilon(y)$ is psd of rank $\le t-1$. Thus, by (2.6), length$(f)\le t-1$. QED

For the full cage $C = C_{n,m}$, we write $\Lambda(n, m) = \Lambda(C_{n,m})$. In the case of $n$-ary quadratic forms, for example, we have $a(n, 2) = l(n, 2) = n(n+1)/2$ and $g(x) = (x-n)(x+n+1)$, so $\Lambda(n, 2) = n = e(n, 2)$. In this case, (4.2) and (4.4) give the best bound for the pythagoras number, since we know that $P(n, 2) = n$. But in general we cannot expect $\Lambda(C)$ to be a sharp upper bound for $P(C)$ since the derivation of $\Lambda(C)$ did not take into account the specific geometric configuration of $C$. A case in point is the (full) cage of binary $m$-ics, $C = C_{2,m}$. For this cage, $a(2, m) = l(2, m) = m + 1$, and $\Lambda(2, m) = (\sqrt{8m+9} - 1)/2$, but of course we know that $P(2, m) = 2$!

In the ternary case $(n = 3)$, the upper bound $\Lambda$ takes a particularly simple form for the full cage. Here $a = a(3, m) = (m+1)(m+2)/2$, so $1 + 8a = (2m+3)^2$, leading to $P(3, m) \leq \Lambda(3, m) = m + 1$. During the Conference, David Leep informed us that, by using quadratic form theory over the field $\mathbf{R}(x)$, he has shown that $P(3, m) \leq \frac{m}{2} + 2$. His proof **[Le]** for this sharper upper bound (in the ternary case) will appear elsewhere. (Unfortunately, neither bound would give Hilbert's impressive result $P(3, 4) = 3$.)

To conclude this section, we shall apply (4.4) to a cage which has proved to be of considerable importance in the literature. This is the cage of $(n_1, n_2; m_1, m_2)$ biforms, i.e., forms of the shape $f = \sum c_\alpha x^\alpha$, where $\alpha = (a_1, \dots, a_{n_1}; b_1, \dots, b_{n_2})$ with $\sum a_i = m_1$ and $\sum b_j = m_2$. These are forms which can be viewed as an $n_1$-ary $m_1$-ic in one set of variables whose "coefficients" are $n_2$-ary $m_2$-ic in another set of variables. The cage $C = C(n_1, n_2; m_1, m_2)$ arising from these biforms is essentially the "direct product" $C_{n_1,m_1} \times C_{n_2,m_2}$. From this and (3.4), we see easily that $A(C) = L(C)$, and that $a(C) = a(n_1, m_1)a(n_2, m_2)$, $e(C) = e(n_1, m_1)e(n_2, m_2)$. Thus, we get from (4.4):

COROLLARY 4.5. *For even $m_1, m_2$, the pythagoras number for $(n_1, n_2; m_1, m_2)$ biforms is at most $(\sqrt{1+8a} - 1)/2$, where $a = \binom{n_1+m_1-1}{n_1-1} \cdot \binom{n_2+m_2-1}{n_2-1}$.*

To get a feeling for this bound, look at the case of *biquadratic* forms: $(m_1, m_2) = (2, 2)$. Here the pythagoras number is at most

$$(\sqrt{1 + 2n_1n_2(n_1+1)(n_2+1)} - 1)/2.$$

Taking $(n_1, n_2; m_1, m_2) = (n, 2; 2, 2)$, for instance, this bound is

$$\frac{\sqrt{1 + 12n(n+1)} - 1}{2} \leq \frac{\sqrt{3}(2n+1) - 1}{2},$$

so the pythagoras number for biquadratic forms of the shape

$$f(x_1, \dots, x_n; y, z)$$

is at most $[\sqrt{3}n + \frac{\sqrt{3}-1}{2}]$, as was announced in our earlier work **[CLR, §7]**. We should note, however, that David Leep **[Le]** has obtained better pythagoras number bounds in this case, and indeed in the case when $(n_1, n_2; m_1, m_2) = (n, 2; 2, m)$.

## §5. Dimension and interior of $F^+(C)$ and $F(C)$

Recall that, for a cage $C \subseteq C_{n,m}$, $F^+(C)$ is the family of psd $n$-ary $m$-ics $f$ with $C(f) \subseteq C$, and $F(C)$ is the subfamily of $F^+(C)$ consisting of sums of squares. These two sets can be made into topological spaces in a natural way, by regarding two forms as "close" if their respective coefficients are close in the real line topology. More precisely, we embed $F^+(C)$ in $\mathbf{R}^l$ $(l = l(C))$ by taking $f = \sum_{\alpha \in L(C)} c_\alpha x^\alpha \in F^+(C)$ to $(c_\alpha) \in \mathbf{R}^l$ and pull back the topology from $\mathbf{R}^l$. Similarly, we embed $F(C)$ in $\mathbf{R}^a$ $(a = a(C))$ by taking $f = \sum_{\alpha \in A(C)} c_\alpha x^\alpha \in F(C)$ (cf. (3.7)(1)) to $(c_\alpha) \in \mathbf{R}^a$ and pull back the topology from $\mathbf{R}^a$. In this way, *we shall view $F^+(C)$ as a cone in $\mathbf{R}^l$, and $F(C)$ as a cone in $\mathbf{R}^a$*. (For more details on this viewpoint, see [**Re**$_2$].)

In this section, we shall determine the "degrees of freedom" for these two cones. This will be important for deriving lower bounds for $P(C)$ in the next section. To talk in more precise language, we shall use the notion of "dimension" for topological spaces. For an exposition of dimension theory (for separable metric spaces[4]), see the book of Hurewicz and Wallman [**HW**]. More pertinent for our present purposes is the following basic fact:

**5.1.** A subset $F \subseteq \mathbf{R}^k$ has dimension $k$ iff $F$ has a nonempty interior, i.e., iff $F$ contains a nonempty open ($k$-dimensional) ball.

For any cage $C \subseteq C_{n,m}$, let $\mathrm{Int}(F^+(C))$ denote the interior of $F^+(C)$ as a subspace of $\mathbf{R}^l$, and let $\mathrm{Int}(F(C))$ denote the interior of $F(C)$ as a subspace of $\mathbf{R}^a$. Recall that the principal form of the case $C$ is the form $f_C := \sum_{\alpha \in E(C)} x^\alpha$. The key result of this section is the following:

THEOREM 5.2. *The principal form $f_C$ lies in both* $\mathrm{Int}(F^+(C))$ *and* $\mathrm{Int}(F(C))$.

In particular, $\mathrm{Int}(F^+(C))$ and $\mathrm{Int}(F(C))$ are both nonempty. It follows from (5.1) that:

COROLLARY 5.3. *For any cage $C \subseteq C_{n,m}$,*
(1) $\dim F^+(C) = l := l(C)$,
(2) $\dim F(C) = a := a(C)$.

PROOF OF THEOREM 5.2. We show first that $f_C \in \mathrm{Int}(F^+(C))$. The argument here exploits the Arithmetic-Geometric Inequality. Observe that this inequality has the following formulation: If $\alpha_i$ *are even lattice points and* $\gamma = \sum \lambda_i \alpha_i$ with $\lambda_i \geq 0$, $\sum \lambda_i = 1$, then $\sum \lambda_i x^{\alpha_i} \geq |x^\gamma|$. Let $\gamma \in L(C)$; say

[4]The spaces in question, $F^+(C)$ and $F(C)$, are, in fact, semialgebraic spaces. For a modern treatment of dimension theory for general semialgebraic spaces, see [**DK**, §8].

$\gamma = \sum_{\alpha\in E(C)} \lambda_\alpha \alpha$, with $\lambda_\alpha \geq 0$, $\sum \lambda_\alpha = 1$. Then

$$|x^\gamma| \leq \sum_{\alpha\in E(C)} \lambda_\alpha x^\alpha \leq \sum_{\alpha\in E(C)} x^\alpha = f_C.$$

Take any form $h(x) = \sum_{\gamma\in L(C)} \varepsilon_\gamma x^\gamma$ with $\sum |\varepsilon_\gamma| \leq 1$. Then

$$h(x) \leq \sum_{\gamma\in L(C)} |\varepsilon_\gamma| \cdot |x^\gamma| \leq \sum_{\gamma\in L(C)} |\varepsilon_\gamma| \cdot f_C \leq f_C,$$

i.e., $f_C - h \in F^+(C)$. This shows that $f_C \in \mathrm{Int}(F^+(C))$.

To show that $f_C \in \mathrm{Int}(F(C))$, take any $\gamma \in A(C)$, say $\gamma = \frac{1}{2}(\alpha_1 + \alpha_2)$, $\alpha_i \in E(C)$. From

$$a^{\alpha_1} + x^{\alpha_2} \pm 2x^\gamma = (x^{\alpha_1/2} \pm x^{\alpha_2/2})^2,$$

we see that $f_C \pm 2x^\gamma$ are both sums of squares. Given any form $h(x) = \sum_{\gamma\in A(C)} 2\varepsilon_\gamma x^\gamma$ with $\sum |\varepsilon_\gamma| \leq 1$, we have then

$$f_C - h = \left(1 - \sum |\varepsilon_\gamma|\right) f_C + \sum |\varepsilon_\gamma| (f_C \pm 2x^\gamma) \in F(C).$$

Thus, $f_C \in \mathrm{Int}(F(C))$. QED

In the special case when $C$ is the full cage $C_{n,m}$, the dimension formulas in (5.3) were first proved by R. M. Robinson **[Ro]**. However, the interior form in $F^+(C)$ and $F(C)$ discovered by Robinson was *not* the principal form; rather, it is the form $g(x) = x_1^m + \cdots + x_n^m$. The fact that $g \in \mathrm{Int}(F(C_{n,m}))$ can be seen directly as follows. Let $\gamma \in L(C_{n,m})$. By a theorem in Hardy, Littlewood, and Pólya **[HLP**, p. 55], $\sum x_i^m \pm x^\gamma$ is a sum of squares. Thus, for any form $h(x) = \sum \varepsilon_\gamma x^\gamma$ with $\sum |\varepsilon_\gamma| \leq 1$, we have

$$g(x) - h(x) = \left(1 - \sum |\varepsilon_\gamma|\right) \sum x_i^m + \sum |\varepsilon_\gamma| \left(\sum x_i^m \pm x^\gamma\right) \in F(C_{n,m}).$$

This shows that $g \in \mathrm{Int}(F(C_{n,m}))$, and hence, also $g \in \mathrm{Int}(F^+(C_{n,m}))$ since $a(C_{n,m}) = l(C_{n,m})$.

In the balance of this section, we shall characterize the forms in $\mathrm{Int}(F^+(C))$, $\mathrm{Int}(F(C))$, and establish an inclusion relation between these two sets. We first deal with $\mathrm{Int}(F^+(C))$.

PROPOSITION 5.4. *For any cage $C \subseteq C_{n,m}$ and $f \in F^+(C)$, the following statements are equivalent:*

(1) $f \in \mathrm{Int}(F^+(C))$;

(2) $f \geq \varepsilon f_C$ *for some* $\varepsilon > 0$.

*In case $C = C_{n,m}$, these are also equivalent to*

(3) *$f$ is a strictly definite form (i.e., $f$ has no nontrivial real zeros).*

PROOF. (1) $\Rightarrow$ (2) is clear. For the converse, assume that $f \geq \varepsilon f_C$, where $\varepsilon > 0$. Since $f_C \in \mathrm{Int}(F^+(C))$, we have $f_C \geq h$ for any form $h = \sum_{\gamma\in L(C)} \varepsilon_\gamma x^\gamma$, where $\{\varepsilon_\gamma\}$ are sufficiently small, and hence, $f \geq \varepsilon f_C \geq$

$\varepsilon h$. This shows that $f \in \operatorname{Int}(F^+(C))$. Now assume $C = C_{n,m}$. If $f \in \operatorname{Int}(F^+(C))$, then $f \geq \varepsilon(x_1^m + \cdots + x_n^m)$ for some $\varepsilon > 0$, so clearly (3) holds. Conversely, assume (3) holds. Then, on the unit sphere $S^{n-1}$, we have $f \geq \varepsilon$ for some $\varepsilon > 0$. For any form $h = \sum \varepsilon_\gamma x^\gamma$ with $\sum |\varepsilon_\gamma| \leq \varepsilon$, and $x \in S^{n-1}$, we have

$$f(x) - h(x) \geq \varepsilon - \sum |\varepsilon_\gamma| \cdot |x^\gamma| \geq \varepsilon - \sum |\varepsilon_\gamma| \geq 0,$$

and so $f \geq h$ everywhere, whence $f \in \operatorname{Int}(F^+(C))$. QED

We shall now characterize the forms in $\operatorname{Int}(F(C))$.

PROPOSITION 5.5. *For any cage $C \subseteq C_{n,m}$, and $f \in F(C)$, the following statements are equivalent:*

(1) $f \in \operatorname{Int}(F(C))$;
(2) $f - \varepsilon f_C \in F(C)$ *for some* $\varepsilon > 0$;
(3) $f$ *admits an* $e \times e$ *nonsingular Gram matrix* ($e = e(C)$).

PROOF. (1) $\Rightarrow$ (2) is clear from the definition of the interior.

(2) $\Rightarrow$ (3) Say $f = \varepsilon f_C + g$, where $\varepsilon > 0$ and $g \in F(C)$. Recall that $I_e$ (the $e \times e$ identity matrix) is the Gram matrix associated with the expression $f_C = \sum_{\beta \in \frac{1}{2}E(C)} (x^\beta)^2$. Let $V$ be a Gram matrix associated with $g$. Then, by Lemma 2.7, $\varepsilon I_e + V$ is a Gram matrix for $f$. Since $V$ is psd, $\varepsilon I_e + V$ is clearly nonsingular.

(3) $\Rightarrow$ (1) Suppose $f(x) = \sum_{\alpha \in A(C)} a_\alpha x^\alpha$ has an $e \times e$ nonsingular Gram matrix $(u_{\beta\beta'})$. Consider a perturbation $f_\varepsilon(x) = f(x) + \sum_{\alpha \in A(C)} \varepsilon_\alpha x^\alpha$ of $f$. For $\{\varepsilon_\alpha\}$ sufficiently small, we shall show that $f_\varepsilon(x)$ also admits a Gram matrix, so (by 2.4)) $f_\varepsilon \in F(C)$. The desired Gram matrix $(v_{\beta\beta'})$ for $f_\varepsilon$ will be obtained by a perturbation of $(u_{\beta\beta'})$, say $v_{\beta\beta'} = u_{\beta\beta'} + \varepsilon_{\beta\beta'}$, where $\varepsilon_{\beta\beta'} = \varepsilon_{\beta'\beta}$. The numbers $v_{\beta\beta'}$ must satisfy

$$\sum_{(\beta,\beta') \in D(\alpha)} v_{\beta\beta'} = a_\alpha + \varepsilon_\alpha \qquad (\forall \alpha \in A(C)),$$

so the $\varepsilon_{\beta\beta'}$ must satisfy

$$\sum_{(\beta,\beta') \in D(\alpha)} \varepsilon_{\beta\beta'} = \varepsilon_\alpha \qquad (\forall \alpha \in A(C)).$$

Clearly, this is satisfied if we choose $\varepsilon_{\beta\beta'} = \varepsilon_\alpha/|D(\alpha)|$, where $\alpha = \beta + \beta'$ (the symmetry condition $\varepsilon_{\beta\beta'} = \varepsilon_{\beta'\beta}$ is automatic for these choices). If the $\varepsilon_\alpha$'s are small, so are the $\varepsilon_{\beta\beta'}$'s. Since $(u_{\beta\beta'})$ is nonsingular, it is positive definite, and so its small perturbations will also be positive definite. Thus, for $\{\varepsilon_\alpha\}$ sufficiently small, $(v_{\beta\beta'})$ will be a (nonsingular) Gram matrix for $f_\varepsilon$, as desired. QED

REMARK. Of course, the Proposition above gives a new proof for $f_C \in \operatorname{Int}(F(C))$. But the proof given earlier for (5.2) was considerably easier.

COROLLARY 5.6. *For any cage* $C \subseteq C_{n,m}$, *we have*

(1) $\mathrm{Int}(F(C)) \subseteq \mathrm{Int}(F^+(C))$;

(2) $\mathrm{Int}(F(C)) = F(C) \cap \mathrm{Int}(F^+(C))$ *iff* $F^+(C) \cap \mathbf{R}^a = F(C)$. *(Here, we identify* $\mathbf{R}^a$ *with* $\mathbf{R} \times \cdots \times \mathbf{R} \times \{0\} \times \cdots \times \{0\}$ *in* $\mathbf{R}^l$).

PROOF. (1) is clear by the characterizations in Propositions (5.4) and (5.5).

(2) Suppose $F^+(C) \cap \mathbf{R}^a = F(C)$. Let $f \in F(C) \cap \mathrm{Int}(F^+(C))$. By (5.4), we have $f \geq \varepsilon f_C$ for some $\varepsilon > 0$. Then $f - \varepsilon f_C \in F^+(C) \cap \mathbf{R}^a = F(C)$, so by (5.5), $f \in \mathrm{Int}(F(C))$. This shows that

$$F(C) \cap \mathrm{Int}(F^+(C)) \subseteq \mathrm{Int}(F(C)),$$

and by (1), this is an equality. Conversely, assume that $F^+(C) \cap \mathbf{R}^a \neq F(C)$. Take a form $f \in F^+(C) \cap \mathbf{R}^a$ such that $f \notin F(C)$. Since $f_C \in \mathrm{Int}(F(C))$ and $f \in \mathbf{R}^a$, we have $f_C + \varepsilon f \in F(C)$ for sufficiently small $\varepsilon$. Thus, the set

$$N := \{\eta \geq 0 \colon f + \eta f_C \in F(C)\}$$

is nonempty. Note that $F(C)$ is a *closed* subspace of $\mathbf{R}^a$ (this can be easily established by an argument similar to that in [**Ro**, p. 268]). Thus, $N$ is a closed set in $\mathbf{R}$. Let $\eta_0$ be the least element of $N$. Since $f \notin F(C)$, $\eta_0 > 0$. Then, by (5.4),

$$g := f + \eta_0 f_C \in \mathrm{Int}(F^+(C)) \cap F(C),$$

but, for $\varepsilon > 0$ (no matter how small),

$$g - \varepsilon f_C = f + (\eta_0 - \varepsilon) f_C \notin F(C)$$

by the choice of $\eta_0$. Thus, $g \notin \mathrm{Int}(F(C))$, and we have a strict containment $F(C) \cap \mathrm{Int}(F^+(C)) \supset \mathrm{Int}(F(C))$. QED

To conclude this section, we would like to mention a result from function theory and dimension theory which will be needed in the next section. This concerns the behavior of dimension under polynomial mappings. While examples such as Peano's space-filling curve show that dimensions may increase under a continuous mapping, it is known that the same cannot happen under *polynomial* maps. Thus, if $H \colon \mathbf{R}^w \to \mathbf{R}^a$ is a polynomial mapping (i.e., the coordinate functions are given by polynomials), then we must have $\dim(\mathrm{im}\, H) \leq w$. In particular, our determination of $\dim F(C)$ in (5.3)(2) has the following consequence.

PROPOSITION 5.7. *If* $H \colon \mathbf{R}^w \to \mathbf{R}^{a(C)}$ *is a polynomial mapping such that* $\mathrm{im}(H) \supseteq F(C)$, *then* $w \geq a(C)$.

Since this result will be crucial to §6, we sketch a direct proof here for the convenience of the reader. Assume, instead, that $w < a := a(C)$. Let $h_i(x_1, \ldots, x_w)$, $1 \leq i \leq a$, be the coordinate functions of $H$. Since the number of polynomials involved is larger than the number of variables, the $h_i$'s cannot be algebraically independent. Thus, there exists a relation

$F(h_1, \ldots, h_a) \equiv 0$ for some nonzero polynomial $F \in \mathbf{R}[t_1, \ldots, t_a]$. But then $\mathrm{im}(H)$ is contained in the real hypersurface defined by $F(t_1, \ldots, t_a) = 0$. This is a contradiction since the hypothesis $\mathrm{im}(H) \supseteq F(C)$ implies that $\mathrm{im}(H)$ contains an open ball in $\mathbf{R}^a$, according to (5.3)(2). (Similar ideas can be used to prove the more general result mentioned in the paragraph preceding the Proposition.)

## §6. Lower bounds for the Pythagoras number

The results of the last section allow us to establish a lower bound for the pythagoras number of a cage.

THEOREM 6.1. *For any cage $C \subseteq C_{n,m}$, let $P = P(C)$, $a = a(C)$, and $e = e(C)$. Then $P^2 - (2e+1)P + 2a \leq 0$. To be more explicit, let*

$$\lambda = \lambda(C) := \frac{2e + 1 - \sqrt{(2e+1)^2 - 8a}}{2}$$

*be the smaller root of the quadratic equation $g(x) := x^2 - (2e+1)x + 2a = 0$. Then $P \geq \lambda \geq a/e$.*

PROOF. We label points in $A(C)$ by $\alpha_1, \ldots, \alpha_a$ and points in $\frac{1}{2}E(C)$ by $\beta_1, \ldots, \beta_e$. All forms $f \in F(C)$ can be written

$$f = \sum_{i=1}^{P} h_i^2 \quad \text{with } h_i(x) = \sum_{j=1}^{e} u_j^{(i)} x^{\beta_j}. \tag{6.2}$$

At first blush, there are $eP$ independent "variables" $u_j^{(i)}$ for $1 \leq j \leq e$ and $1 \leq i \leq P$. But the Gram-Schmidt Lemma (2.12) shows that we can write $f$ in such a way that $u_j^{(i)} = 0$ for $i > j$. This "cuts a corner" off the block of variables, and, since $P \leq e$, leaves a total of $w := eP - \frac{1}{2}P(P-1)$ independent variables. Now define a mapping $H\colon \mathbf{R}^w \to \mathbf{R}^a$ so that the $l$th coordinate function of $H$ is the coefficient of $x^{\alpha_l}$ in the expression for $f$ in (6.2), namely:

$$\sum_{\beta_j + \beta_k = \alpha_l} \sum_{i=1}^{P} v_j^{(i)} v_k^{(i)}.$$

Viewing $\{v_j^{(i)} : i \leq j\}$ as independent variables, we see that $H$ is a polynomial (in fact quadratic) mapping. Hence, by Proposition 5.7, $eP - \frac{1}{2}P(P-1) \geq a$. This means $P^2 - (2e+1)P + 2a \leq 0$, so $P \leq \lambda$, the smaller root of $g(x) := x^2 - (2e+1)x + 2a = 0$ given in (6.1). If we carry out the argument above without the "clipped corner", we will get a weaker inequality $eP \geq a$, so we must have $\lambda \geq a/e$. Alternatively, note that $g(a/e) = a(a-e)/e^2 \geq 0$. Since $a/e \leq (e+1)/2 \leq (2e+1)/2$, this implies that $a/e \leq \lambda$. QED

REMARK 6.3. The Upper Bound Theorem (4.4) and the Lower Bound Theorem (6.1) together show that $\lambda(C) \leq \Lambda(C)$. This, of course, can also be verified directly: From $\Lambda^2 + \Lambda = 2a$, we get $g(\Lambda) = 2\Lambda(\Lambda - e)$. Since $0 < \Lambda \leq e$, we have $g(\Lambda) \leq 0$, so $\Lambda \geq \lambda$.

Applying (4.4) and (6.1) to the full cage $C_{n,m}$, we obtain an explicit description of the asymptotic behavior of the pythagoras numbers $P(n, m)$ where we fix (the degree) $m$ and let (the number of variables) $n$ go to infinity:

THEOREM 6.4. *There exist positive constants* $\gamma_1(m)$ *and* $\gamma_2(m)$ *such that* $\gamma_1(m)n^{m/2} \leq P(n, m) \leq \gamma_2(m)n^{m/2}$.

PROOF. Let $a = a(n, m) = \binom{n+m-1}{m}$, and $e = e(n, m) = \binom{n+\frac{m}{2}-1}{\frac{m}{2}}$. By (4.4), we have

$$P(n, m) \leq \frac{\sqrt{1+8a}-1}{2} \leq \sqrt{2a} = \sqrt{\frac{2}{m!}} \cdot \sqrt{n(n+1)\cdots(n+m-1)},$$

and by (6.1), we have

$$P(n, m) \geq a/e = \frac{(m/2)!}{m!}\left(n+\frac{m}{2}\right)\cdots(n+m-1).$$

From these bounds, it is clear that

$$\overline{\lim_{n\to\infty}}\, n^{-m/2}P(n, m) \leq \sqrt{\frac{2}{m!}} \quad \text{and} \quad \underline{\lim_{n\to\infty}}\, n^{-m/2}P(n, m) \geq \frac{(m/2)!}{m!},$$

from which the theorem follows. QED

If $m = 2$, (6.4) becomes $\gamma_1(2)n \leq P(n, 2) \leq \gamma_2(2)n$. But of course, we know that $P(n, 2) = n$. If $m = 4$, it is not difficult to work out the bounds $\Lambda(n, 4)$ and $\lambda(n, 4)$, as $a(n, 4) = n(n+1)(n+2)(n+3)/24$ and $e(n, 4) = n(n+1)/2$. Omitting the deals, one computes that

(6.5)

$$\Lambda(n, 4) = \frac{1}{2\sqrt{3}}(n^2+3n+1) - \frac{1}{2} + o(1) \qquad \left(\frac{1}{2\sqrt{3}} \approx 0.289\right),$$

$$\lambda(n, 4) = \left(\frac{1}{2}-\frac{1}{\sqrt{6}}\right)n^2 + \frac{1}{2}n + \frac{1}{2}\left(1+\frac{1}{\sqrt{6}}\right) + o(1) \qquad \left(\frac{1}{2}-\frac{1}{\sqrt{6}} \approx 0.092\right).$$

These formulas, therefore, give asymptotic estimates of the pythagoras number for quartic forms in terms of the number of variables.

Recall that, if $C$, $C'$ are two cages such that $C' \subseteq C$, then $P(C') \leq P(C)$. Somewhat surprisingly, the lower bound $\lambda$-function is *not* monotone with respect to the inclusion relation. For instance, let $C'$ be the cage discussed in Example 3.9, where $e(C') = 4$, $a(C') = 10$, and hence, $\lambda(C') = 4$. But for $C = C_{3,6} \supseteq C'$, we have $e(3, 6) = 10$, $a(3, 6) = 28$, and hence,

$$\lambda(C_{3,6}) = (21 - \sqrt{217})/2 \approx \pi < \lambda(C').$$



In this particular example, we get no additional information since $P(3, 6) \geq 4$ and $P(3, 6) \geq \pi$ are the same statement. But in general, $P(C) \geq \lambda(C')$ may be sharper than $P(C) \geq \lambda(C)$. This suggests the following definition: $\overline{\lambda}(C) = \sup\{\lambda(C')\}$, where the supremum is taken over *all* cages $C' \subseteq C$. It follows that, for any cage $C$, $P(C) \geq \overline{\lambda}(C)$. To show that this is indeed a strengthening of (6.1), let us compare, for instance, $\lambda(n, 4)$ and $\overline{\lambda}(n, 4)$ for the cage $C = C_{n,4}$.

Let $C'$ be the "subcage" of $C_{n,4}$ with the corners clipped off, so $E(C')$ consists of all permutations of $(2, 2, 0, \dots, 0)$. It is easy to see that $e(C') = e(n, 4) - n$, and $a(C') = a(n, 4) - n^2$. A routine computation shows that

$$\lambda(C') = \left(\frac{1}{2}-\frac{1}{\sqrt{6}}\right)n^2 + \left(\frac{3}{\sqrt{6}}-\frac{1}{2}\right)n + \left(\frac{1}{2}-\frac{1}{\sqrt{6}}\right) + o(1).$$

Comparing this with the second formula in (6.5), we get

$$\lambda(C') - \lambda(C) = \left(\frac{3}{\sqrt{6}}-1\right)n - \frac{3}{2\sqrt{6}} + o(1) \qquad \left(\frac{3}{\sqrt{6}} - 1 \approx 0.225\right),$$

so in fact $\lim_{n\to\infty}(\overline{\lambda}(n, 4) - \lambda(n, 4)) = \infty$.

The the convenience of the reader, we compile the following Table 1 for $\lambda(n, m)$ and $\Lambda(n, m)$ for $n \leq 5$ and $m \leq 8$, rounding off to three decimal places:

| $n$ | $m$ | $e(n, m)$ | $a(n, m)$ | $\lambda(n, m)$ | $\Lambda(N, m)$ |
|---|---|---|---|---|---|
| 2 | 2 | 2 | 3 | 2 | 2 |
| 2 | 4 | 3 | 5 | 2 | 2.702 |
| 2 | 6 | 4 | 7 | 2 | 3.275 |
| 2 | 8 | 5 | 9 | 2 | 3.772 |
| 3 | 2 | 3 | 6 | 3 | 3 |
| 3 | 4 | 6 | 15 | 3 | 5 |
| 3 | 6 | 10 | 28 | 3.135 | 7 |
| 3 | 8 | 15 | 45 | 3.242 | 9 |
| 4 | 2 | 4 | 10 | 4 | 4 |
| 4 | 4 | 10 | 35 | 4.156 | 7.882 |
| 4 | 6 | 20 | 84 | 4.618 | 12.471 |
| 4 | 8 | 35 | 165 | 5 | 17.673 |
| 5 | 2 | 5 | 15 | 5 | 5 |
| 5 | 4 | 15 | 70 | 5.488 | 11.343 |
| 5 | 6 | 35 | 210 | 6.513 | 20 |
| 5 | 8 | 70 | 495 | 7.411 | 30.968 |

Thus, $P(3, 6) \in \{4, 5, 6, 7\}$, $P(4, 4) \in \{5, 6, 7\}$, and $P(5, 4) \in \{6, 7, 8, 9, 10, 11\}$, etc.

**Acknowledgment**

The authors would like to thank Robin Sahner for her assistance with the LATEX work on the diagrams and charts in this paper.

## REFERENCES

[BS] T. Bier and U. Schwardmann, *Räume normierter Bilinearformen und Clifford-strukturen*, Math. Zeit. **180** (1982), 203–215.

[CEP] J. W. S. Cassels, W. J. Ellison, and A. Pfister, *On sums of squares and on elliptic curves over function fields*, J. Number Theory **3** (1971), 125–149.

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

[$CL_1$] M. D. Choi and T. Y. Lam, *An old question of Hilbert*, Proceedings of Conference on Quadratic Forms (G. Orzech, ed.), Queen's Papers in Pure and Appl. Math., vol. 46, Queen's Univ. Press, Kingston, Ontario, 1977, pp. 385–405.

[$CL_2$] ——, *Extreme positive semidefinite forms*, Math. Ann. **231** (1977), 1–18.

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

[DK] H. Delfs and M. Knebusch, *Semialgebraic topology over a real closed field*. II, Math. Zeit. **178** (1981), 175–213.

[Hi] D. Hilbert, *Über die Darstellung definiter Formen als Summe von Formen-quadsraten*, Math. Ann. **32** (1888), 342–350.

[HLP] G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities*, Second Ed., Cambridge Univ. Press, London and New York, 1962.

[HW] W. Hurewicz and H. Wallman, *Dimension theory*, Princeton University Press, Princeton, NJ, 1941.

[La] T. Y. Lam, *The algebraic theory of quadratic forms*, printing with revisions, W. A. Benjamin, Inc./Addison-Wesley, Reading, MA, 1980.

[Le] D. Leep, private communication.

[Pf] A. Pfister, *Zur Darstellung definiter Funktionen als Summe von Quadraten*, Invent. Math. **4** (1967), 229–237.

[Ra] A. R. Rajwade, *Squares*, London Math. Soc. Lecture Note Ser., vol. 171, Cambridge Univ. Press, London and New York, 1993.

[$Re_1$] B. Reznick, *Extremal psd forms with few terms*, Duke Math. J. **45** (1978), 363–374.

[$Re_2$] ——, *Sums of even powers of linear forms*, Mem. Amer. Math. Soc., vol. 96, Amer. Math. Soc., Providence, RI, 1992.

[Ro] R. M. Robinson, *Some definite polynomials which are not sums of squares of real polynomials*, Selected Questions of Algebra and Logic, Acad. Sci. USSR, 1973, pp. 264–282.

[Sw] R. G. Swan, *Hilbert's theorem on positive ternary quartics*, these Proceedings.

[Yu] S. Yuzvinsky, *Orthogonal pairings of Euclidean spaces*, Michigan Math. J. **28** (1981), 131–145.

UNIVERSITY OF TORONTO, TORONTO M5S 1A1, CANADA

UNIVERSITY OF CALIFORNIA, BERKELEY, CALIFORNIA 94720

UNIVERSITY OF ILLINOIS, URBANA, ILLINOIS 61801