Waring's problem for binary forms

U. Helmke

Department of Mathematics, University of Regensburg, 8400 Regensburg, Germany

Communicated by M.-F. Roy
Received 19 January 1990
Revised 10 September 1991

Abstract


We consider the classical problem of decomposing a binary form of degree $d$ into a sum of $d$th powers of linear forms. The problem is solved for arbitrary binary forms with coefficients in a real or algebraically closed field. Our analysis is based on continued fractions and Padé approximation theory. A connection with partial realizations in control theory is made.

1. Introduction

Let $K$ denote a field of characteristic zero. For any $d \in \mathbb{N}$ let $V_d$ denote the $K$-vector space of binary forms

$$\phi(X; Y) = \sum_{j=0}^{d} \left( \binom{d}{j} h_{j+1} X^{d-j} Y^j \right),$$

$h_{j+1} \in K$, of degree $d$. Waring's problem for binary forms is the following:

**Problem 1.** (i) Given a form $\phi \in V_d$, can $\phi$ be written as a sum of $d$th powers of linear forms

$$\phi(X, Y) = \sum_{i=1}^{s} (a_i X + b_i Y)^d$$

with coefficients $a_i, b_i \in K$?
(ii) Given a form $\phi \in V_d$, can $\phi$ be written as a weighted sum of $d$th powers

$$\phi(X, Y) = \sum_{j=1}^{s(\phi)} c_j (a_j X + b_j Y)^d$$

(1.2b)

with coefficients $a_j, b_j, c_j \in K$, $c_j \neq 0$?

Of course, once such a representation (1.2a) or (1.2b) for a form $\phi$ is established, one would also like to know the minimal number $s$ of summands arising in (1.2a), (1.2b). Let $s(\phi)$ and $s_\infty(\phi)$ respectively denote the minimal number of summands, necessary in the decomposition (1.2a) and (1.2b) respectively. We set $s(\phi) = \infty$ or $s_\infty(\phi) = \infty$, if no decomposition (1.2a) or (1.2b) exists.

**Problem 2.** (i) For any binary form $\phi \in V_d$, compute $s(\phi)$ and parametrize all minimal decompositions

$$\phi(X, Y) = \sum_{j=1}^{s(\phi)} (a_j X + b_j Y)^d.$$  

(1.3a)

(ii) For $\phi \in V_d$, compute $s_\infty(\phi)$ and parametrize all minimal decompositions

$$\phi(X, Y) = \sum_{j=1}^{s_\infty(\phi)} c_j (a_j X + b_j Y)^d.$$  

(1.3b)

Finally we also mention the classical task (Problem 3) of finding suitable canonical decompositions ('canonical forms') of a binary form. These might however have a more general form than by a sum (1.3a) of $d$th powers.

This problem as well as the search for invariants and covariants connected with the canonical forms will not be addressed in this paper. Instead, we focus on Problems 1 and 2.

The above problems have a long history. In the classical literature, Problems 1–3 were studied and partially solved for binary forms with complex coefficients. Important contributions are due to Sylvester [18, 19] and Gundelfinger (1886). Sylvester determines $s(\phi)$ for a generic class of binary forms of odd degree with complex coefficients, but he also considers certain nongeneric classes of binary forms. An effective algorithm to determine the canonical form of a binary form of odd degree has been given by Dür [2]. For a modern account of the theory of forms of odd degree we refer to Lascoux [14]. For forms of even degree the task of finding a canonical form, as attempted by Sylvester [18, 19], is more complicated; see Kung's paper [12]. Gundelfinger expresses the minimal number of summands $s(\phi)$ in terms of the Gundelfinger covariants, valid for a generic class of binary forms. For a systematic modern exposition of the invariant theory of binary forms, based on the so-called umbral calculus, we refer to the survey paper of Kung and Rota [13].
Despite of the classical nature of Waring's problem, a complete solution is still missing. In particular, for nongeneric choices of binary forms, a formula for the minimal numbers $s(\phi)$ and $s_w(\phi)$ is apparently unknown. In this paper we use methods from control theory, viz. partial realization theory, to solve Waring's problem for binary forms with coefficients in an algebraically or real closed field $K$. Our solution seems to be new even in the classical case of binary forms with complex coefficients and is based on an analysis of the Hankel matrix associated with any binary form.

A similar approach for forms with coefficients in an algebraically closed field has been developed by Reichstein [23]. However, his solution is not as general and precise as ours. Also a formula for the minimal number of summands is not given.

Given a binary form

$$\phi(X, Y) = \sum_{j=0}^{d} \binom{d}{j} h_{j+1} X^{d-j} Y^j$$

with coefficients $h_{j+1} \in K$, the Hankel of $\phi$ is defined by

$$\mathcal{H}(\phi) = \begin{bmatrix} h_1 & \cdots & h_n \\ \vdots & \ddots & \vdots \\ h_n & \cdots & h_{2n-1} \end{bmatrix}$$

for $d = 2n - 2$ even and by

$$\mathcal{H}(\phi) = \begin{bmatrix} h_1 & \cdots & h_n & h_{n+1} \\ \vdots & \ddots & \vdots & \vdots \\ h_n & \cdots & h_{2n-1} & h_{2n} \end{bmatrix}$$

if $d = 2n - 1$ is odd. Our main results are the following theorems:

**Theorem A.** Let $K$ be a field of characteristic zero. Every binary form $\phi \in V_d$ is a weighted sum (1.3b) of $d$th powers of linear forms with

$$\text{rank } \mathcal{H}(\phi) \leq s(\phi) \leq d + 1.$$ 

**Theorem B.** Let $K = \bar{K}$ be an algebraically closed field of characteristic zero. Then $s(\phi) = s_w(\phi)$ for all $\phi \in V_d$.

(a) Let $d = 2n - 2$. $s(\phi)$ is either equal to rank $\mathcal{H}(\phi)$ or equal to $n + \text{corank } \mathcal{H}(\phi)$. $s(\phi) = \text{rank } \mathcal{H}(\phi)$ holds generically. Up to multiples of the coefficients $(a_j, b_j)$ in (1.3a) by $d$th roots of unity, there exists a unique decomposition (1.3a) if $s(\phi) < n$. For $n \leq s(\phi) = 2n - 1$, there exists a $(2(s(\phi) - n) + 1)$-parameter family of solutions of (1.3a).

(b) Let $d = 2n - 1$. $s(\phi)$ is either equal to rank $\mathcal{H}(\phi)$ or equal to $n + 1 + \text{corank } \mathcal{H}(\phi)$. $s(\phi) = \text{rank } \mathcal{H}(\phi)$ holds generically. Up to multiples of the co-
coefficients \((a_j, b_j)\) in (1.3a) by \(d\)th roots of unity, there exists a unique decomposition (1.3a) if \(s(\phi) \leq n\). If \(n < s(\phi) \leq 2n\), there exists a \(2(s(\phi) - n)\)-parameter family of solutions.

**Theorem C.** Let \(K\) be a real closed field and \(d = 2n - 2, n \in \mathbb{N}\). Waring’s problem (1.3a) is solvable over \(K\) if and only if the Hankel \(\mathcal{H}(\phi)\) is positive semidefinite. Furthermore, \(s(\phi) = s_w(\phi) = \text{rank } \mathcal{H}(\phi)\). For rank \(\mathcal{H}(\phi) < n\) there exists—up to multiples of \((a_j, b_j)\) by \(\pm 1\)—a unique solution for (1.3a) while for rank \(\mathcal{H}(\phi) = n\), a semialgebraic one-parameter family of solutions exists.

As already mentioned, our main technical tool for the study of Waring’s Problems 1 and 2 comes from control theory and is known as partial realization theory. A closely related theory is that of Padé approximation. Padé approximation is equivalent to continued fraction expansions of formal power series and has been developed by Wronski, Cauchy and Jacobi. A more recent reference is [1]. Partial realization theory has been developed by Kalman [9, 10] and is nowadays a standard and well-established tool in control theory with effective solution algorithms available. Given any rational function \(g(s) \in K(s), g(\infty) = 0\), there is associated with it a ‘realization’ \((A, b, c) \in K^{n \times n} \times K^n \times K^{1 \times n}\) such that

\[
g(s) = c(sI - A)^{-1}b
\]  

(1.4)

holds. If \(g(s) = \sum_{i=1}^{N} g_s s^{-i}\) is a finite Laurent polynomial, then \((A, b, c)\) is called an \(N\)th order partial realization of \(g(s)\) if

\[
c(sI - A)^{-1}b = \sum_{i=1}^{N} g_s s^{-i} + O(s^{-(N+1)})
\]  

(1.5)

holds. Of course, such (partial) realizations can only be uniquely determined up to transformations \((A, b, c) \mapsto (SAS^{-1}, Sb, cS^{-1})\), which all leave (1.4) or (1.5) invariant.

The existence of such partial realizations being guaranteed, our crucial but simple observation (Lemma 4.1) is that every binary form \(\phi(X, Y)\) can be written as

\[
\phi(X, Y) = c(XI + YA)^d b.
\]  

(1.6)

(This can be easily generalized to forms in more than two variables.) Thus we establish a link between Waring’s problem and canonical forms for the similarity action \((A, b, c) \mapsto (SAS^{-1}, Sb, cS^{-1})\), and thus in particular with the Jordan canonical form. Therefore, the questions of finding canonical forms for binary forms and that of canonical forms for the similarity action \((A, b, c) \mapsto (SAS^{-1}, Sb, cS^{-1})\) are closely related and in fact equivalent topics.
For example, by putting $A$ into the Jordan canonical form, the Sylvester canonical form for forms of odd degree is obtained. Similarly, the computational task of finding the Sylvester canonical form of a binary form of odd degree reduces to the computation of a partial realization $(A, b, c)$, with $A$ in Jordan canonical form. We will not address these issues any further.

We proceed as follows. In Section 2 we introduce the Hankel associated with every binary form. The natural $GL(2, K)$-action on the vector space of binary forms induces an action on rectangular Hankel matrices and an important Lemma 2.2 allows us to reduce the analysis to the case of regular Hankels. Section 3 gives a brief survey of partial realization theory. The main technical result is Theorem 3.4 from which Theorems A–C are deduced in Section 4.

2. Hankel matrices

Let $K$ be a field of characteristic zero. For integers $M, N \geq 1$ let $\text{Hank}(M \times N)$ denote the $K$-vector space of all rectangular $M \times N$ Hankel matrices

$$\mathcal{H} = \begin{bmatrix} h_1 & \cdots & h_N \\ \vdots & \ddots & \vdots \\ h_M & \cdots & h_{M+N-1} \end{bmatrix}$$

with entries $h_{i+j-1} \in K$. Given a binary form

$$\phi(X, Y) = \sum_{j=0}^{d} \binom{d}{j} h_{j+1} X^{d-j} Y^j,$$

the $Hankel$ of $\phi$ is defined by

$$\mathcal{H}(\phi) = \begin{bmatrix} h_1 & \cdots & h_n \\ \vdots & \ddots & \vdots \\ h_n & \cdots & h_{2n-1} \end{bmatrix} \in \text{Hank}(n \times n)$$

if $d = 2n - 2$ is even and by

$$\mathcal{H}(\phi) = \begin{bmatrix} h_1 & \cdots & h_n & h_{n+1} \\ \vdots & \ddots & \vdots & \vdots \\ h_n & \cdots & h_{2n-1} & h_{2n} \end{bmatrix} \in \text{Hank}(n \times (n + 1))$$

if $d = 2n - 1$ is odd. Clearly $\phi \mapsto \mathcal{H}(\phi)$ gives a $K$-linear isomorphism of $V_d$ onto $\text{Hank}(n \times n)$, $\text{Hank}(n \times (n + 1))$ respectively.

For any $d \in \mathbb{N}$, the general linear group $GL(2, K)$ of invertible $2 \times 2$ matrices

$$g = \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix}$$
defines a right action on $V_d$ via
\[ \tau : \text{GL}(2, K) \times V_d \to V_d, \quad (g, \phi) \mapsto \phi \cdot g, \] (2.5)
where
\[ (\phi \cdot g)(X, Y) := \phi(g_{11}X + g_{12}Y, g_{21}X + g_{22}Y). \] (2.6)

This group action of $\text{GL}(2)$ or $\text{SL}(2)$ plays a central role in the classical theory of binary forms; see e.g. [20] and [21].

The induced linear representation is the $d$th symmetric power
\[ \tau_{d+1} : \text{GL}(2, K) \to \text{GL}(d + 1, K) \] (2.7)
defined by
\[ \tau_{d+1}(g)\phi(X, Y) = \phi(g_{11}X + g_{21}Y, g_{12}X + g_{22}Y) \] (2.8)
and is irreducible. It is easy to give an explicit formula for the entries of the $(d + 1) \times (d + 1)$ matrix $\tau_{d+1}(g)$ (see e.g. [4]):
\[ \tau_{d+1}(g)_{ij} = \sum_{p=0}^{d} \binom{i}{p} \binom{d-j}{i-p} g_{11}^{i-p} g_{12}^{i-j} g_{21}^{j-p} g_{22}^{d-j-p}. \] (2.9)

Let
\[ \sigma := \text{diag}\left(\left(\begin{smallmatrix} d \\ 0 \end{smallmatrix}\right), \left(\begin{smallmatrix} d \\ 1 \end{smallmatrix}\right), \ldots, \left(\begin{smallmatrix} d \\ d \end{smallmatrix}\right)\right). \] (2.10)

Using (2.9), it is easily seen that
\[ \sigma \cdot \tau_{d+1}(g)^T \cdot \sigma^{-1} = \tau_{d+1}(g^T) \] (2.11)
holds for all $g \in \text{GL}(2, K)$. Since every $g \in \text{GL}(2, K)$ can be written as a product of elementary transformations
\[ g(a) = \begin{bmatrix} a & 0 \\ 0 & 1 \end{bmatrix}, \quad a \neq 0, \] (2.12a)
\[ w = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \] (2.12b)
\[ g' = \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix}, \quad t \in K, \] (2.12c)
it suffices to determine $\tau_{d+1}(g)$ for these special matrices. Applying (2.9) yields
\[ \tau_{d+1}(g(a)) = \text{diag}(a^d, a^{d-1}, \ldots, a, 1), \quad a \neq 0, \quad (2.13a) \]

\[ \tau_{d+1}(w) = \begin{bmatrix} 0 & \cdots & 1 \\ \vdots \\ 1 & \cdots & 0 \end{bmatrix}_{(d+1) \times (d+1)} \quad (2.13b) \]

\[ \tau_{d+1}(g') = \begin{pmatrix} 0 & \cdots & (d-1) & d \\ 0 & \cdots & \frac{1}{d} t^{d-1} & \frac{1}{d} t^d \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \frac{1}{d} t & \frac{1}{d} t^d \end{pmatrix} \quad (2.13c) \]

The following lemma is an immediate consequence of Theorem 4.1 in [4] and (2.11).

**Lemma 2.1.** For every \( g \in \text{GL}(2, K), \phi \in V_d \),

\[ \mathcal{H}(\phi \cdot g) = \tau_n(g)^T \mathcal{H}(\phi) \tau_n(g) \quad (2.14a) \]

if \( d = 2n - 2 \) is even and

\[ \mathcal{H}(\phi \cdot g) - \tau_n(g)^T \mathcal{H}(\phi) \tau_{n+1}(g) \quad (2.14b) \]

if \( d = 2n - 1 \) is odd. Here \( \mathcal{H}(\phi) \) is the Hankel defined by (2.3) and \( \phi \cdot g \) is defined by (2.6). \( \square \)

The lemma shows that the \( \text{GL}(2, K) \)-action on \( V_d \) leaves the rank of the associated Hankel matrices \( \mathcal{H}(\phi) \) invariant. Obviously \( s(\phi) \) and \( s_w(\phi) \) are also invariant under \( \text{GL}(2, K) \). For a proof of the following lemma we refer to [17].

**Lemma 2.2.** For \( \phi \in V_d \) let \( r = \text{rank} \mathcal{H}(\phi) \). For almost every \( g \in \text{GL}(2, K) \) (in the Zariski-topology of \( \text{GL}(2, K) \)), the \( r \times r \)-principal minor of \( \mathcal{H}(\phi \cdot g) \) is nonzero. \( \square \)

We call a Hankel matrix \( \mathcal{H} \) regular if the rank \( \mathcal{H} \times \text{rank} \mathcal{H} \) principal minor of \( \mathcal{H} \) is nonzero. Using Lemma 2.1 and 2.2 we can restrict ourselves to the case of regular Hankels.

**Remark 2.3.** In classical invariant theory, see e.g. [13], the determinant of the Hankel matrix \( \mathcal{H}(\phi) \) is called the catalecticant and thus is defined only for forms of even degree. Lemma 2.1 implies the well-known result that the catalecticant of a form \( \phi \in V_d \) of even degree is a \( \text{GL}(2, K) \)-invariant. More generally, the minors of Hankel matrices are Schur functions which have been systematically studied by Lascoux [15].
3. Partial realization theory

Our main technical tool for studying Waring's problem comes from control theory and is known as partial realization theory. Partial realization theory was developed mainly by Kalman [9, 10], and has its roots in the early work of Kronecker [11] and Frobenius. There is a close connection with Padé approximation theory, moment problems and interpolation theory. For further references we refer to [3, 6, 7, 9, 16, 17].

**Definition 3.1.** Let \((h_1, \ldots, h_N) \in K^N, N \in \mathbb{N} \cup \{\infty\}\). A rational function \(g(s) \in K(s), g(\infty) = 0,\) is called a *Padé approximation* of \((h_1, \ldots, h_N)\) if

\[
g(s) = \sum_{j=1}^{N} h_j s^{-j} + O(s^{-N-1}) .
\]

(3.1)

The following state-space version of Definition 3.1 is standard in control theory:

**Definition 3.1'.** Let \(h = (h_1, \ldots, h_N), N \in \mathbb{N} \cup \{\infty\}\). A triple \((A, b, c) \in K^{n \times n} \times K^{n \times 1} \times K^{1 \times n}\) is called an *n-dimensional partial realization* of \(h\) if

\[
h_j = cA^{j-1}b \quad \text{for } j = 1, \ldots, N .
\]

(3.1')

There is a simple correspondence between both types of realizations.

If \((A, b, c)\) is a partial realization of the sequence \(h\), then \(g(s) = c(sI - A)^{-1}b\) is a Padé approximation of \(h\). Conversely, if a rational function \(g(s) \in K(s), g(\infty) = 0,\) is a Padé approximant for \(h\), then any \((A, b, c)\) with

\[
c(sI - A)^{-1}b = g(s)
\]

(3.2)

is a partial realization for \(h\). Moreover, if \(g(s) \in K(s), g(\infty) = 0,\) is given, then there exists \((A, b, c)\) satisfying (3.2) (see Lemma 3.2).

Let \(g(s) = p(s)/q(s) \in K(s), p\) and \(q\) coprime, be a Padé approximant for \(h = (h_1, \ldots, h_N)\). Then

\[
\deg(g) = \max(\deg p, \deg q)
\]

(3.3)

is called the *order* or *degree* of the approximant. An *n-dimensional partial realization* \((A, b, c)\) of \((h_1, \ldots, h_N)\) is called *irreducible* if \((A, b, c)\) satisfies the generic rank conditions

\[
\rank(b, Ab, \ldots, A^{n-1}b) = n ,
\]

(3.4a)

\[
\rank(\begin{pmatrix} c \\
   cA \\
   \vdots \\
   cA^{n-1} \
\end{pmatrix}) = n .
\]

(3.4b)
Conditions (3.4a) and (3.4b) are equivalent to the conditions that the linear system

\[
\begin{align*}
\dot{x}(t) &= Ax(t) + bu(t) \\
y(t) &= cx(t)
\end{align*}
\]  

is controllable and observable; Kalman (1969).

**Lemma 3.2.** Let \( h = (h_1, \ldots, h_N) \in K^N, N \in \mathbb{N}, \) and a monic polynomial \( q(s) \in K(s) \) of degree \( N \) be given. There exists a realization \((A, b, c)\) for \( h \) such that \( q(s) \) is the characteristic polynomial of \( A \). Furthermore, there exists an irreducible realization \((A, b, c)\) for \( h \) of dimension \( \equiv N \) such that the characteristic polynomial of \( A \) divides \( q(s) \).

**Proof.** Let \( \tilde{A} := (h_1, \ldots, h_N), \quad \tilde{b} = (1, 0, \ldots, 0)^T \) and let

\[
\tilde{A} = \begin{bmatrix}
0 & \cdots & 0 & -q_0 \\
1 & \cdots & 0 & -q_1 \\
\vdots & \ddots & \ddots & \vdots \\
0 & \cdots & 1 & -q_{N-1}
\end{bmatrix}
\]  

be the companion matrix of the polynomial \( q(s) = \sum_{j=0}^{N-1} q_j s^j, q_N = 1 \). \((\tilde{A}, \tilde{b}, \tilde{c})\) is an \( N \)-dimensional (partial) realization for \( h \). Let \( V = \bigcap_{j=0}^{N-1} \ker \tilde{A}^j \) and let \( W \subset K^N \) denote a complementary subspace of \( V \), so that \( K^N = V \oplus W \). \( V \) is \( \tilde{A} \)-invariant and therefore, up to a change of basis in \( K^N \), \((\tilde{A}, \tilde{b}, \tilde{c})\) is of the form

\[
\begin{bmatrix}
A_{11} & A_{12} \\
0 & A
\end{bmatrix}, \quad \begin{bmatrix}
b_1 \\
b
\end{bmatrix}, \quad \begin{bmatrix}
0 \\
c
\end{bmatrix}
\]  

with \((A, b, c)\) an irreducible realization for \( h \). Obviously the characteristic polynomial of \( A \) divides that of \( \tilde{A} \), which is \( q(s) \). \( \square \)

For infinite sequences \( h = (h_j \mid j \in \mathbb{N}) \) the realization question is solved by Kronecker's theorem [10]. Let \( H(h) = (h_{i+j-1})_{i,j \in \mathbb{N}} \) denote the infinite Hankel matrix associated with \( h \). See [10] for a proof of the following lemma:

**Lemma 3.3.** An infinite sequence \( h \in K^\mathbb{N} \) has a finite-dimensional realization \((A, b, c)\) if and only if \( \text{rank} H(h) < \infty \). In that case, \( \text{rank} H(h) \) is equal to the dimension of any irreducible realization \((A, b, c)\) of \( h \). \( \square \)

The partial realization theory for finite sequences \( h = (h_1, \ldots, h_N) \) is in many
ways more subtle than the realization theory for infinite sequences. It is easy to see that any finite sequence \( h \in K^N \) has an infinite number of irreducible partial realizations, whose dimensions grow unbounded. In particular, there is no analogue of Lemma 3.3 for finite sequences. For any sequence \( h = (h_1, \ldots, h_N) \in K^N \), \( N \in \mathbb{N} \), let

\[
H(h) = \begin{bmatrix}
h_1 & \cdots & h_n \\
\vdots & & \vdots \\
h_n & \cdots & h_{2n-1}
\end{bmatrix} \in \text{Hank}(n \times n)
\]  

(3.8a)

if \( N = 2n - 1 \) is odd and

\[
H(h) = \begin{bmatrix}
h_1 & \cdots & h_n & h_{n+1} \\
\vdots & & \vdots & \vdots \\
h_n & \cdots & h_{2n-1} & h_{2n}
\end{bmatrix} \in \text{Hank}(n \times (n + 1))
\]  

(3.8b)

if \( N = 2n \) is even. For simplicity we assume in the sequel that \( H(h) \) is regular, i.e. the \( r \times r \)-principal minor of \( H(h) \) is nonzero, where \( r = \text{rank} \; H(h) \).

Let \( r = \text{rank} \; H(h) \). Consider the polynomials defined by

\[
p_H(s) := \det \begin{bmatrix}
0 & h_1 & h_1 s + h_2 & \cdots & h_1 s^{r-1} + \cdots + h_r \\
h_1 & h_2 & \cdots & \vdots & \vdots \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
h_r & h_{r+1} & h_{r+2} & \cdots & h_{2r}
\end{bmatrix},
\]  

(3.9a)

\[
q_H(s) := \det \begin{bmatrix}
s & \cdots & s^r \\
h_1 & h_2 & \cdots & h_{r+1} \\
\vdots & \vdots & \ddots & \vdots \\
h_r & h_{r+1} & \cdots & h_{2r}
\end{bmatrix}.
\]  

(3.9b)

(For \( r = n \) and \( N = 2n - 1 \) odd, \( p_H \) and \( q_H \) depend on an additional parameter \( h_{2n} \in K \).) It has been shown by Jacobi [8], see also [16], that the \( n \times n \) resp. \( n \times (n + 1) \) Hankel of the rational function \( g_H = p_H/q_H \in K(s) \) coincides with \( H(h) \). Therefore, \( (p_H, q_H) \) are coprime and define a realization \( p_H/q_H \) of \( h \) of minimal degree rank \( H(h) \) (this uses the fact that \( H(h) \) is regular!).

**Theorem 3.4.** For \( h \in K^N \) let \( H(h) \) be the regular Hankel defined by (3.8).

(i) Any Padé approximant \( g(s) \) satisfies \( \text{deg}(g) > \text{rank} \; H(h) \).

(ii) If \( N \) is odd and \( H(h) \) invertible, the set of rank \( H(h) \)-order Padé-approximants for \( h \) is given by the one-parameter family \( g_H = p_H/q_H \) in \( h_{2n} \in K \), where \( p_H, q_H \) are defined by (3.9).

(iii) If \( N \) is even or if \( H(h) \) is not full rank, \( g_H = p_H/q_H \) defined by (3.9) is the uniquely determined Padé approximant of order rank \( H(h) \).

(iv) There does not exist a Padé approximant \( g \) of \( h \) whose degree satisfies

\[
\text{rank} \; H(h) < \text{deg}(g) < N + 1 - \text{rank} \; H(h)
\]  

(3.10)
(v) The set of Padé approximants \( g \in K(s) \) of \( h \) of order \( N + 1 - \text{rank} \ H(h) \) depends on \( N + 2 - 2 \text{rank} \ H(h) \) parameters.

(vi) There always exists a Padé approximant \( g = p/q \in K(s) \) for \( h \) of order \( N + 1 - \text{rank} \ H(h) \), such that the polynomial \( q \) is separable.

**Proof.** (i) follows immediately from Kronecker’s theorem, see Lemma 3.3. For proofs of (ii), (iii) we refer to [10, 16]. To prove the other statements we use the continued-fraction representation of rational functions. Without loss of generality let \( \text{corank} \ H(h) \geq 1 \). Let \( g = p/q \) denote the unique Padé approximant for \( h \) of degree \( r = \text{rank} \ H(h) \). By expanding \( p/q \) into a continued fraction we have

\[
\frac{p(s)}{q(s)} = \frac{\beta_0}{\alpha_1(s)} - \frac{\beta_1}{\alpha_2(s)} - \cdots - \frac{\beta_{k-1}}{\alpha_k(s)} - \frac{\beta_k}{\alpha_{k+1}(s)} - \cdots
\]

where \( \alpha_i(s) \in K[s] \) are monic polynomials with \( \sum_{i=1}^k \deg \alpha_i = r \) and \( \beta_i \in K - \{0\} \), \( i = 0, \ldots, k - 1 \), are nonzero constants. \( \alpha_i, i = 1, \ldots, k \), are called the atoms of \( p/q \) and \( \beta_i, i = 0, \ldots, k - 1 \), are called the phases of \( p/q \). By [16] (see also Theorem 4.4 in [5]), the continued-fraction expansion of any other Padé approximant \( \hat{g}(s) \) for \( h \) of degree \( \deg(\hat{g}) > \text{rank} \ H(h) \) must have the form

\[
\hat{g}(s) = \frac{\beta_0}{\alpha_1} - \frac{\beta_1}{\alpha_2} - \cdots - \frac{\beta_{k-1}}{\alpha_k} - \frac{\beta_k}{\alpha_{k+1}} - \cdots
\]

with \( l > k \) and \( \hat{g} \) has the same first \( k \) atoms \( \alpha_i \) and phases \( \beta_{l-i} \) as \( p/q \) for \( i \leq k \). For any rational function \( f \in K(s) \), \( f(\infty) = 0 \), let \( (h_i(f) \mid i \in \mathbb{N}) \) be defined by

\[
f(s) = \sum_{i=1}^\infty h_i(f)s^{-i}.
\]

Since \( p/q \) and \( \hat{g} \) are Padé approximant for \( h \in K^N \)

\[
h_i(\hat{g}) = h_i(p/q) \quad \text{for } i = 1, \ldots, N.
\]
By (4.13) in [5]

\[ h_i(\hat{g}) = h_i(p/q) \quad \text{for} \quad 1 \leq i < 2r + \deg \alpha_{k+1} \]  

and

\[ h_i(\hat{g}) \neq h_i(p/q) \]  

for \( i = 2r + \deg \alpha_{k+1} \).

Thus

\[ \deg \alpha_{k+1} > N - 2r \]  

which implies

\[
\deg(\hat{g}) = \sum_{i=1}^{l} \deg \alpha_i \geq \sum_{i=1}^{k} \deg \alpha_i + \deg \alpha_{k+1} \\
\geq r + N - 2r + 1 = N - 1
\]

which proves (iv). To prove (v), we observe, that for \( \deg(\hat{g}) = N - r + 1 \), then \( \hat{g} \) is obtained from \( p/q \), by adding one atom \((\alpha_{k+1}, \beta_i)\) of degree \( N - 2r + 1 \). Thus \( \hat{g} \) depends on \( N - 2r + 2 \) parameters, which proves (v). Let \( g = p/q \in K(s), \ g(\infty) = 0, \) be a Padé approximant for \( h \) of order \( d := N - \text{rank} H(h) + 1 \). Let \( \hat{h} = (\hat{h}_1, \ldots, \hat{h}_{2d-1}) \), where \( \hat{h}_i := h_i(g) \) is defined via (3.14). By Kronecker's theorem, the Hankel matrix

\[
H := H_d(p/q) = \begin{bmatrix}
\hat{h}_1 & \cdots & \hat{h}_{2d-1} \\
\vdots & \ddots & \vdots \\
\hat{h}_{2d-1} & \cdots & \hat{h}_{2d-1}
\end{bmatrix}
\]

is invertible. By (ii), \( \tilde{g}_\hat{h} = p_{\hat{h}}/q_{\hat{h}} \) with \( p_{\hat{h}}, q_{\hat{h}} \) defined by (3.9) (and \( r \) replaced by \( d \)), are realizations for \( (\hat{h}_1, \ldots, \hat{h}_{2d-1}) \). From (3.9),

\[
g_{\hat{h}} = \frac{p_{\hat{h}}}{q + \xi a}
\]

in the parameter \( \xi = \hat{h}_{2d} \in K \). Here \( a \in K[s] \) is the unique solution of the Bezout identity

\[
ap + bq = 1
\]

with \( \deg a < \deg q \). (vi) now follows from the fact that for arbitrary coprime polynomials \((a, q)\), \( \deg a < \deg q \), there exist infinitely many \( \xi \in K \) such that \( q + \xi a \in K[s] \) is separable. □
Remark 3.5. A corresponding result holds for partial realizations \((A, b, c)\), if everywhere in Theorem 3.4 ‘Padé-approximant’ is replaced by ‘irreducible realization \((A, b, c)\)’.

4. Proofs of the main results

The proofs of our main results rest on the following simple lemma. Let \(K\) denote a field of characteristic zero.

Lemma 4.1. Let \(\phi(X, Y) = \sum_{j=0}^{d} \binom{d}{j} h_{j+1} X^{d-j} Y^j\) with \(h_{j+1} \in K, j = 0, \ldots, d\). Then for every partial realization \((A, b, c)\) of \(h = (h_1, \ldots, h_{d+1})\),

\[
\phi(X, Y) = c(XI + YA)^d b.
\]

Proof. Trivial, since

\[
c(XI + YA)^d b = \sum_{j=0}^{d} \binom{d}{j} cA^jbX^{d-j}Y^j
\]

\[
= \sum_{j=0}^{d} \binom{d}{j} h_{j+1} X^{d-j} Y^j
\]

\[
= \phi(X, Y).
\]

Proof of Theorem A. Let \(\phi(X, Y) = \sum_{j=0}^{d} \binom{d}{j} h_{j+1} X^{d-j} Y^j\). Choose a polynomial

\[
q(s) = \prod_{i=0}^{d} (s - a_i)
\]

with \(a_i \in K, a_i \neq a_j\) for \(i \neq j\). By Lemma 3.2, there exists a realization \((A, b, c) \in K^{(d+1) \times (d+1)} \times K^{(d+1) \times 1} \times K^{1 \times (d+1)}\) for \(h = (h_1, \ldots, h_{d+1})\) with

\[
A = \text{diag}(a_0, \ldots, a_d),
\]

\[
b = (b_0, \ldots, b_d)^T, \quad c = (c_0, \ldots, c_d).
\]

By Lemma 4.1

\[
\phi(X, Y) = c(XI + YA)^d b = \sum_{j=0}^{d} c_j b_j (X + a_j Y)^d.
\]

Thus \(s_\infty(\phi) \leq d + 1\). For \(s < \text{rank } H(\phi)\) suppose

\[
\phi(X, Y) = \sum_{j=1}^{s} \gamma_j (a_j X + \beta_j Y)^d.
\]
Using the $\text{GL}(2, K)$-action on $V_d$ we may assume that $\alpha_j \neq 0$ for $j = 1, \ldots, s$. Thus

$$\phi(X, Y) = \sum_{j=1}^{s} c_j (X + a_j Y)^d$$

with $c_j = \gamma_j \alpha_j^d$, $a_j = \beta_j / \alpha_j$, $j = 1, \ldots, s$. It follows that

$$A = \text{diag}(a_1, \ldots, a_s),$$
$$b = (b_1, \ldots, b_s)^T,$$
$$c = (c_1, \ldots, c_s) \in K^{1 \times r},$$

is a realization of $\mathcal{H}(\phi)$ of dimension $s < \text{rank } \mathcal{H}(\phi)$, contradicting Theorem 3.4(i). Thus $\delta_w(\phi) \geq \text{rank } \mathcal{H}(\phi)$, which completes the proof. □

To prove Theorem B, we now assume that $K = \bar{K}$ is algebraically closed. For any $\phi \in V_d$, let $\mathcal{H}(\phi)$ denote the Hankel matrix, defined by (2.3). Thus $\mathcal{H}(\phi) \in \text{Hank}(n \times n)$ for $d = 2n - 2$ and $\mathcal{H}(\phi) \in \text{Hank}(n \times (n + 1))$ for $d = 2n - 1$, $n \in \mathbb{N}$. Let $\text{Hank}(n \times n)^{\text{sep}}$ (resp. $\text{Hank}(n \times (n + 1))^{\text{sep}}$) denote the subset of all $n \times n$ (resp. $n \times (n + 1)$) Hankel matrices $H$ which are $\text{GL}(2, K)$ equivalent to a regular Hankel with an irreducible realization $(A, b, c)$ of dimension $\text{rank } H$, such that the characteristic polynomial of $A$ is separable, i.e. splits into rank $H$ distinct linear factors (over $\bar{K}$). Consider the subsets

$$V_{d}^{\text{sep}} := \{ \phi \in V_d \mid \mathcal{H}(\phi) \in \text{Hank}(n \times n)^{\text{sep}} \},$$

$$V_{d}^{\text{sep}} := \{ \phi \in V_d \mid \mathcal{H}(\phi) \in \text{Hank}(n \times (n + 1))^{\text{sep}} \}.$$

for $d = 2n - 2$, $d = 2n - 1$ odd respectively. By Lemmas 2.1 and 2.2, $V_{d}^{\text{sep}}$ is a constructible Zariski-dense subset of $V_d \cong K^{d \times 1}$. Using Theorem 3.4(ii), (iii) and (vi) the binary forms in $V_{d}^{\text{sep}}$ have the following characterization:

$$\phi \in V_{d}^{\text{sep}} \iff \mathcal{H}(\phi) \text{ is } \text{GL}(2, K)\text{-equivalent to a Hankel } H' \text{ with an irreducible realization } (A, b, c) \text{ of dimension } r = \text{rank } \mathcal{H}(\phi):$$

$$A = \text{diag}(a_1, \ldots, a_r), \quad a_i \neq a_j \in K,$$
$$b = (b_1, \ldots, b_r)^T \in K^{r \times 1},$$
$$c = (c_1, \ldots, c_r) \in K^{1 \times r},$$
$$c_i b_j \neq 0 \quad \text{for } i = 1, \ldots, r.$$

Thus, using Lemma 4.1 and Lemma 2.1, we obtain the following lemma:

**Lemma 4.2.** $\phi \in V_{d}^{\text{sep}}$ if and only if...
\[
\phi(X, Y) = \sum_{j=1}^{r} (a_j X + b_j Y)^d
\]  
(4.5)

with \( r = \text{rank } \mathcal{H}(\phi) \) and pairwise linear independent vectors \( \left( \frac{a_j}{b_j} \right) \in K^2, \ j = 1, \ldots, r. \)

In particular,
\[
\phi \in V^*_{d} \Leftrightarrow s(\phi) = s_w(\phi) = \text{rank } \mathcal{H}(\phi).
\]  
(4.6)

Now consider the nongeneric case, where \( \phi \in V_d - V^*_{d} \). Using Lemmas 2.1 and 2.2, Theorem 3.4(iv), (vi) implies
\[
\phi \in V_d - V^*_{d} \Rightarrow s(\phi) = s_w(\phi) = d + 2 - \text{rank } \mathcal{H}(\phi).
\]  
(4.7)

This completes the proof of Theorem B. 

To prove Theorem C, let \( K \) be a real closed field and \( d = 2n - 2 \) even. To prove the necessity of the condition, suppose
\[
\phi(X, Y) = \sum_{j=1}^{s} (a_j X + b_j Y)^d,
\]
with \( s \in \mathbb{N} \) and \( a_j, b_j \in K \). Without loss of generality, we can assume that \( a_j \neq 0, \ j = 1, \ldots, s \). Define for \( j = 1, \ldots, s \):
\[
\lambda_j := b_j / a_j, \quad \beta_j := a_j^{n-1}.
\]

Then
\[
A := \text{diag}(\lambda_1, \ldots, \lambda_s), \quad b = (\beta_1, \ldots, \beta_s)^T, \quad c = (\beta_1, \ldots, \beta_s).
\]  
(4.8a, 4.8b)

is a realization of \( \mathcal{H}(\phi) \). Thus
\[
\mathcal{H}(\phi) = [b, Ab, \ldots, A^{r-1}b]^T \cdot [b, Ab, \ldots, A^{r-1}b] \succeq 0,
\]
from which the necessity follows. To prove the sufficiency, suppose that \( \mathcal{H}(\phi) \succeq 0 \) with \( \text{rank } \mathcal{H}(\phi) = r \leq n \). By Lemmas 2.1 and 2.2 we can assume without loss of generality that \( \mathcal{H}(\phi) \) is regular. Then there exists a symmetric, irreducible realization \( (A, b, c) \in K^{r \times r} \times K^{r \times 1} \times K^{1 \times r} \) such that
\[
A = A^T, \quad b = c^T.
\]  
(4.9)

see [22].
Using an appropriate orthogonal transformation $S \in O(r, K)$, $(A, b, c)$ is similar to $(F, g, h) = (SAS^T, Sb, cs^T)$ with

$$F = \text{diag}(\lambda_1, \ldots, \lambda_r),$$

$$\lambda_1 < \cdots < \lambda_r, \quad \lambda_i \in K,$$

$$g^T = h = (\gamma_1, \ldots, \gamma_r), \quad \gamma_i > 0.$$  \hfill (4.10a, 4.10b, 4.10c)

Thus, by Lemma 4.1,

$$\phi(X, Y) = \sum_{j=1}^r \gamma_j^2 (X + \lambda_j Y)^{2n-2} = \sum_{j=1}^r (a_j X + b_j Y)^{2n-2}$$

with $a_j, b_j \in K$ and $a_j^{-1} = \gamma_j, b_j = \lambda_j a_j, j = 1, \ldots, r$. This proves the sufficiency and $s_w(\phi) \simeq s(\phi) \simeq r$. By Theorem A also $s_w(\phi) \simeq r$ and the result follows, using Theorem 3.4.  \hfill $\square$

**Acknowledgment**

The author thanks Professor Alain Lascoux for his helpful remarks on this paper.

**References**

Waring's problem for binary forms