

Rank and Optimal Computation of Generic Tensors*

V. Strassen

*Institut für Angewandte Mathematik
Universität Zürich
Rämistrasse 74
CH-8032 Zürich, Switzerland*

Dedicated to Alexander M. Ostrowski
on the occasion of his ninetieth birthday.

Submitted by Walter Gautschi

ABSTRACT

The typical rank (= maximal border rank) of tensors of a given size and the set of optimal bilinear computations of typical tensors of a given rank are investigated. For the size $(n, n, 3)$ with n odd, the complement of the set of tensors of maximal border rank is a hypersurface. Its equation is given.

INTRODUCTION

An important aspect of A. Ostrowski's pioneering paper "On two problems in abstract algebra connected with Horner's rule" [27] lies in a judicious choice of the cost function for defining computational complexity. The nonscalar cost measure (Ostrowski measure), although somewhat artificial, has proved useful in connection with various methods for obtaining lower bounds (substitution method, degree method; see e.g. [7]) as well as for studying the complexity of matrix multiplication and other bilinear problems.

For such problems a further simplification is crucial, that of restricting to bilinear computations. The loss in efficiency (at most a factor 2) is well paid off by the simplicity of the resulting definition of complexity: Let A, B, W be

*This work was supported in part through funds provided by the National Science Foundation under Grant No. MCS 74-12997 through a subcontract from MIT to the University of Washington.

finite dimensional vector spaces over some field k ,

$$f: A \times B \rightarrow W$$

a bilinear map. The bilinear complexity (or rank) $R(f)$ is the least $r \in \mathbb{N}$ such that there are $u_1, \dots, u_r \in A^*$, $v_1, \dots, v_r \in B^*$, and $w_1, \dots, w_r \in W$ such that

$$\forall a \in A, b \in B, \quad f(a, b) = \sum_{\rho=1}^r u_{\rho}(a) v_{\rho}(b) w_{\rho}. \quad (1.1)$$

If one shifts emphasis to the dual spaces $U = A^*$, $V = B^*$ and represents f by its structural tensor, one gets a more symmetrical definition: Let U, V, W be finite dimensional k -vector spaces, $t \in U \otimes V \otimes W$. The rank $R(t)$ of t is the least $r \in \mathbb{N}$ such that there are $u_1, \dots, u_r \in U$, $v_1, \dots, v_r \in V$, $w_1, \dots, w_r \in W$ such that

$$t = \sum_{\rho=1}^r u_{\rho} \otimes v_{\rho} \otimes w_{\rho}. \quad (1.2)$$

(See Gastinel [12], Strassen [30, 31], Fiduccia [11], Brockett and Dobkin [9], Borodin and Munro [7], and Lafon [25].)

During the past ten years much work has been done to calculate (or estimate from below) the rank of interesting tensors. Let us only mention the determination of the rank of 2-slice tensors ($\dim W = 2$) in terms of their Kronecker normal form for algebraically closed fields by Grigoryev [13] and Ja'Ja [22] (see also Dobkin [10]), and the results on the structural tensors of associative algebras. (See [2] for a survey.) In spite of this, it is clear that present knowledge and techniques are very incomplete.

On the other hand, relatively good bounds are known for the maximum value of the rank of tensors of a given size or "shape" (m, n, q) . Let

$$R(m, n, q) = \max \{ R(t) : t \in U \otimes V \otimes W \},$$

where $\dim U = m$, $\dim V = n$, $\dim W = q$. Then

$$R(m, n, q) \geq \frac{mnq}{m+n+q-2} \quad (1.3)$$

[8, 19, 13] and

$$R(m, n, q) \leq \lfloor q/2 \rfloor n + m, \quad (1.4)$$

assuming w.l.o.g. $m \leq n$ [4].

For what follows it is convenient to let the groundfield k be algebraically closed. Fix U, V, W of dimensions m, n, q respectively. $U \otimes V \otimes W$ is a k -vector space, and therefore also an affine space in the sense of algebraic geometry. Let us define the border rank $\underline{R}(t)$ of a tensor $t \in U \otimes V \otimes W$ by

$$\forall r \quad (\underline{R}(t) \leq r \Leftrightarrow t \in \text{Zariski closure of } \{t' : R(t') \leq r\}). \quad (1.5)$$

Thus \underline{R} is the largest Zariski lower semicontinuous function $\leq R$. (By a result of Alder [1] this is equivalent to the definition of border rank by Bini, Capovani, Lotti, and Romani [6, 5].) Let us denote by $\underline{R}(m, n, q)$ the maximal border rank of tensors $t \in U \otimes V \otimes W$. Then $\underline{R}(m, n, q)$ is also the rank of almost all tensors of $U \otimes V \otimes W$ [i.e. of all tensors outside some lower dimensional subvariety of $U \otimes V \otimes W$; we call $\underline{R}(m, n, q)$ the typical rank for the shape (m, n, q)]. Moreover, if a tensor has algebraically independent coefficients with respect to certain bases of U, V, W , then its rank equals $\underline{R}(m, n, q)$.

Equation (1.3) remains true for \underline{R} [8, 13]. Since $\underline{R} \leq R$, we also have (1.4) for \underline{R} . In the case $m = n$ Atkinson and Lloyd [3] improve this to

$$\underline{R}(n, n, q) \leq \lceil q/2 \rceil n.$$

So for cubic shapes one has

$$\frac{n^2}{3} \leq \underline{R}(n, n, n) \leq \frac{n^2}{2} + n.$$

In the present paper we use linearization via tangential maps (Jacobi criterion) to estimate \underline{R} . One consequence of our results is

$$\underline{R}(m, n, q) \sim \frac{mnq}{m+n+q-2} \quad (1.6)$$

as $m, n, q \rightarrow \infty$ [Corollary 3.6 in combination with Proposition 2.3 and the trivial $\underline{R}(m, n, q) \leq mn$].

Of particular interest are the shapes (m, n, q) such that $\underline{R}(m, n, q) = mnq/(m+n+q-2)$. (Then of course $mnq/(m+n+q-2)$ must be an integer.) We call them perfect. Examples of perfect shapes are $(n, n, n+2)$ for $n \not\equiv 2 \pmod{3}$ and $(n-1, n, n)$ for $n \equiv 0 \pmod{3}$ (Corollary 3.10). Perfect shapes are relatively dense (Corollary 3.11).

Apart from the typical rank of tensors of a given shape, we also study the set of optimal (bilinear) computations of typical tensors of rank r . More precisely, let $X_r \subset U \otimes V \otimes W$ be the closure of the set of tensors of rank $\leq r$.

Then X_r is an irreducible closed subvariety of $U \otimes V \otimes W$. Almost all $t \in X_r$ have rank r and (up to trivial equivalence) exactly one optimal computation, provided that r does not come too close to $\underline{R}(m, n, q)$ (Corollary 3.7; a similar statement holds for q even). (The sets of optimal computation of various special tensors have been studied by Lafon [24], Winograd [32], Kruskal [23], and de Groote [14–17].)

For the shape $(n, n, 3)$ with n odd it turns out that $X_{(3n-1)/2}$ is a hypersurface of $U \otimes V \otimes W$. (In particular, for this shape $\underline{R}(m, n, q) = (3n+1)/2 > (3n-1)/2 = \lceil mnq/(m+n+q-2) \rceil$.) Its equation is exhibited (Theorem 4.6). Thus by evaluating a polynomial one can decide whether any given $(n, n, 3)$ tensor has maximal border rank or not. As an application we determine the border rank of the structural tensor of an arbitrary sl_2 module.

The major development of the paper is self-contained apart from a detailed motivation of the notion of rank and some classical algebraic geometry, which can be found in [28], [29], [26], [18].

Most of Section 4 was written while the author was visiting the Computer Science Department of the University of Washington in Seattle. Its pleasant and stimulating environment is gratefully acknowledged.

T. Lickteig (Universität Konstanz) has independently applied the Jacobian criterion. Moreover he has announced a proof of

$$\underline{R}(n, n, n) = \left\lceil \frac{n^3}{3n-2} \right\rceil \quad (n \neq 3)$$

based on the perfectness results of the present paper.

2. FIBRES OF THE COMPUTATION MAP

Throughout this paper k denotes an algebraically closed field. Let m, n, q be positive integers, and U, V, W be k -vector spaces of dimensions m, n, q respectively. The set

$$S = \{u \otimes v \otimes w : u \in U, v \in V, w \in W\} \subset U \otimes V \otimes W \quad (2.1)$$

of “triads” is an irreducible closed subvariety of dimension $m+n+q-2$ of the vector space $U \otimes V \otimes W$. Moreover it is smooth except at 0. S is called the

Segre variety. Given a positive integer r , we have a morphism of (affine) varieties

$$\begin{aligned} \varphi_r: S^r &\rightarrow U \otimes V \otimes W, \\ (u_\rho \otimes v_\rho \otimes w_\rho)_{\rho \leq r} &\mapsto \sum_{\rho=1}^r u_\rho \otimes v_\rho \otimes w_\rho. \end{aligned} \quad (2.2)$$

We interpret elements of S^r as bilinear computations of length r . φ_r assigns to any such computation the tensor it computes. We therefore call φ_r the computation morphism. By the definition of rank we have

$$\text{im } \varphi_r = \{t \in U \otimes V \otimes W: R(t) \leq r\} \quad (2.3)$$

Let $X_r \subset U \otimes V \otimes W$ denote the closure of $\text{im } \varphi_r$. By the definition of border rank we have

$$X_r = \{t \in U \otimes V \otimes W: \underline{R}(t) \leq r\}. \quad (2.4)$$

X_r is closed and irreducible, and

$$\dim X_r \leq r(m + n + q - 2). \quad (2.5)$$

Almost all $t \in X_r$ have rank $\leq r$. Moreover $X_{r-1} \subset X_r$.

PROPOSITION 2.1. *If $X_{r-1} \neq U \otimes V \otimes W$ then*

$$X_{r-1} \subsetneq X_r$$

and almost all $t \in X_r$ have rank r .

Proof. By way of contradiction assume

$$X_{r-1} = X_r.$$

Then

$$\text{im } \varphi_{r-1} + S \subset X_r = X_{r-1}.$$

Since adding a tensor is continuous and X_{r-1} is closed, we have

$$X_{r-1} + S \subset X_{r-1},$$

and by induction on ν

$$X_{r-1} + \{t: R(t) \leq \nu\} = X_{r-1} + \underbrace{S + \cdots + S}_{\nu \text{ times}} \subset X_{r-1}.$$

But $\{t: R(t) \leq mn\} = U \otimes V \otimes W$, so

$$U \otimes V \otimes W \subset X_{r-1},$$

a contradiction. ■

Now let

$$R(m, n, q) = \max\{R(t) : t \in U \otimes V \otimes W\} = \min\{r : \text{im } \varphi_r = U \otimes V \otimes W\} \quad (2.6)$$

denote the maximal rank, and

$$\underline{R}(m, n, q) = \max\{\underline{R}(t) : t \in U \otimes V \otimes W\} = \min\{r : X_r = U \otimes V \otimes W\} \quad (2.7)$$

the typical rank (or maximal border rank) of tensors of shape (m, n, q) . [By Proposition 2.1 almost all $t \in U \otimes V \otimes W$ have rank $\underline{R}(m, n, q)$, so the name “typical rank” is justified.] The formula (2.5) immediately implies

$$R(m, n, q) \geq \underline{R}(m, n, q) \geq \left\lceil \frac{mnq}{m+n+q-2} \right\rceil. \quad (2.8)$$

In this paper we shall be interested in $\underline{R}(m, n, q)$ rather than in $R(m, n, q)$. We shall see that the second inequality in (2.8) is often, but not always, an equality. Let us look at the more informative inequality (2.5). It can be improved in a trivial way:

$$\dim X_r \leq \min\{r(m+n+q-2), mnq\}. \quad (2.9)$$

This leads to the following definition.

DEFINITION 2.2. $r \geq 1$ is small for (m, n, q) iff

$$\dim X_r = r(m + n + q - 2).$$

r is large for (m, n, q) iff

$$\dim X_r = mnq.$$

Observe that if r is small then $r \leq \lfloor mnq/(m + n + q - 2) \rfloor$ and if r is large then $r \geq \lceil mnq/(m + n + q - 2) \rceil$. (In fact r is large iff φ_r is dominant iff $r \geq \underline{R}(m, n, q)$.)

PROPOSITION 2.3.

(1) We have

$$\begin{aligned} r \text{ small, } s \leq r &\Rightarrow s \text{ small,} \\ r \text{ large, } s \geq r &\Rightarrow s \text{ large.} \end{aligned}$$

(2) Let $(\tilde{m}, \tilde{n}, \tilde{q}) \leq (m, n, q)$ componentwise. Then

$$\begin{aligned} r \text{ small for } (\tilde{m}, \tilde{n}, \tilde{q}) &\Rightarrow r \text{ small for } (m, n, q), \\ r \text{ large for } (m, n, q) &\Rightarrow r \text{ large for } (\tilde{m}, \tilde{n}, \tilde{q}). \end{aligned}$$

Proof. We first make the following observation: Let $t \in U \otimes V \otimes W$, $R(t) = r$, and let

$$t = \sum_{\rho=1}^r u_{\rho} \otimes v_{\rho} \otimes w_{\rho}$$

be an (optimal) computation. Then $u_1 \otimes v_1, \dots, u_r \otimes v_r$ are linearly independent. [Otherwise we would have e.g.

$$u_r \otimes v_r = \sum_{\rho=1}^{r-1} \lambda_{\rho} u_{\rho} \otimes v_{\rho}$$

with $\lambda_{\rho} \in k$. Then

$$t = \sum_{\rho=1}^{r-1} u_{\rho} \otimes v_{\rho} \otimes (w_{\rho} + \lambda_{\rho} w_r),$$

contradicting $R(t) = r$.]

The first part of the proposition being trivial, we prove the second part. W.l.o.g. $(\tilde{m}, \tilde{n}, \tilde{q}) = (m, n, q-1)$. Let (U, V, W) have shape (m, n, q) , and let $\dim \tilde{W} = q-1$. Choose a surjective linear map $\alpha: W \rightarrow \tilde{W}$. Then

$$\beta: U \otimes V \otimes W \rightarrow U \otimes V \otimes \tilde{W},$$

$\beta = \text{id} \otimes \text{id} \otimes \alpha$ is also surjective. Let \tilde{S} be the Segre variety with respect to U, V, \tilde{W} , and $\tilde{\varphi}_r$ the corresponding computation morphism. [See (2.2).] We have the following commutative diagram:

$$\begin{array}{ccc} S^r & \xrightarrow{\varphi_r} & U \otimes V \otimes W \\ \downarrow \beta^r & & \downarrow \beta \\ \tilde{S}^r & \xrightarrow{\tilde{\varphi}_r} & U \otimes V \otimes \tilde{W} \end{array} \quad (2.10)$$

Now if r is large for (m, n, q) , then φ_r is dominant, thus also $\beta\varphi_r$, thus also $\tilde{\varphi}_r\beta^r$, and thus $\tilde{\varphi}_r$. Therefore r is large for $(m, n, q-1)$.

Let r be small for $(m, n, q-1)$. From (2.10) we get the commutative diagram

$$\begin{array}{ccc} S^r & \xrightarrow{\varphi_r} & X_r \\ \downarrow \beta^r & & \downarrow \beta \\ \tilde{S}^r & \xrightarrow{\tilde{\varphi}_r} & \tilde{X}_r \end{array} \quad (2.11)$$

where \tilde{X}_r is the closure of $\text{im } \tilde{\varphi}_r$. There is a dense open set $\tilde{Y} \subset \tilde{X}_r$ such that for $\tilde{t} \in \tilde{Y}$ we have

$$R(\tilde{t}) = r, \quad \tilde{\varphi}_r^{-1}\{\tilde{t}\} \text{ finite.} \quad (2.12)$$

Let $Y = \beta^{-1}(\tilde{Y})$. Since $(\beta^r)^{-1}\tilde{\varphi}_r^{-1}(\tilde{Y}) \neq \emptyset$, we have $Y \neq \emptyset$. It suffices to prove that $\varphi_r^{-1}\{t\}$ is finite for $t \in Y$. In view of (2.12) it is enough to show that

$$\beta^r: \varphi_r^{-1}\{t\} \rightarrow \tilde{\varphi}_r^{-1}\{\tilde{t}\}$$

is injective, where $\tilde{t} = \beta(t) \in \tilde{Y}$. So let $z, z' \in \varphi_r^{-1}\{t\}$, $\beta^r(z) = \beta^r(z')$. Then if

$$z = (u_\rho \otimes v_\rho \otimes w_\rho)_{\rho \leq r},$$

we have

$$z' = (u_\rho \otimes v_\rho \otimes w'_\rho)_{\rho \leq r}.$$

[Observe that $R(\tilde{t}) = r$ implies $\beta(u_\rho \otimes v_\rho \otimes w_\rho) \neq 0$ for all ρ .] Thus we get

$$\sum_{\rho \leq r} u_\rho \otimes v_\rho \otimes w_\rho = t = \sum_{\rho \leq r} u_\rho \otimes v_\rho \otimes w'_\rho.$$

But $u_1 \otimes v_1, \dots, u_r \otimes v_r$ are linearly independent by the remark at the beginning of the proof, so $w_\rho = w'_\rho$ for all ρ and therefore $z = z'$. ■

The next proposition is a direct consequence of some standard results of algebraic geometry.

PROPOSITION 2.4.

(1) Let r be small for (m, n, q) . Then there is a number d such that almost all $t \in X_r$ have rank r and exactly d optimal computations.

(2) Let $r = \underline{R}(m, n, q)$. Then almost all $t \in U \otimes V \otimes W$ have rank r and a set of optimal computations, which is a closed subvariety of S^r of pure dimension $r(m + n + q - 2) - mnq$.

The main result of this section (Theorem 2.7) is a refinement of the first part of the previous proposition. We need an auxiliary result.

LEMMA 2.5. Let $\dim U = m$, $\dim V = n$, $r \leq (m - 1)(n - 1)$. For almost all $((u_\rho, v_\rho))_{\rho \leq r} \in (U \times V)^r$ the following is true: Whenever $\lambda_1, \dots, \lambda_r \in k$, $u \in U$, $v \in V$ such that

$$\sum_{\rho=1}^r \lambda_\rho u_\rho \otimes v_\rho = u \otimes v, \quad (2.13)$$

then $u \otimes v = \lambda_\rho u_\rho \otimes v_\rho$ for some ρ .

Proof. W.l.o.g. $m \leq n$. For $r \leq m$ it suffices to choose u_1, \dots, u_r and v_1, \dots, v_r as linearly independent vectors. So let $r > m \geq 1$. Arguing inductively, it is enough to show that for almost all $u_1, \dots, u_r, v_1, \dots, v_r$ (2.13) is impossible unless at least one $\lambda_\sigma = 0$. Now consider

$$Z \subset U' \times V' \times k' \times U \times V,$$

where $(u_1, \dots, u_r, v_1, \dots, v_r, \lambda_1, \dots, \lambda_r, u, v) \in Z$ iff

$$u_1, \dots, u_m \text{ are linearly independent,}$$

$$\lambda_1 \cdots \lambda_r \neq 0, \quad \sum_{\rho=1}^r \lambda_\rho u_\rho \otimes v_\rho = u \otimes v.$$

Z is a locally closed subvariety of $U^r \times V^r \times k^r \times U \times V$. The projection

$$p: U^r \times V^m \times V^{r-m} \times k^r \times U \times V \rightarrow U^r \times V^{r-m} \times k^r \times U \times V$$

induces an isomorphism of Z onto a nonempty open subset of $U^r \times V^{r-m} \times k^r \times U \times V$. Thus Z is irreducible and

$$\dim Z = mr + n(r - m) + r + m + n \leq mr + nr + 1$$

by the assumption of the lemma. Let

$$f: Z \rightarrow U^r \times V^r$$

by the restriction of the projection

$$U^r \times V^r \times k^r \times U \times V \rightarrow U^r \times V^r.$$

The nonempty fibres of f are at least two dimensional, since $\lambda_1, \dots, \lambda_r, u, v$ may be changed into $\alpha\beta\lambda_1, \dots, \alpha\beta\lambda_r, \alpha u, \beta v$ (where $\alpha, \beta \in k, \alpha\beta \neq 0$) without changing the value of f . Therefore $f(Z)$ is an irreducible constructible subset of $U^r \times V^r$ of dimension $< mr + nr$, i.e.

$$\overline{f(Z)} \subsetneq U^r \times V^r.$$

Now if $(u_1, \dots, u_r, v_1, \dots, v_r) \notin \overline{f(Z)}$, then (2.13) is impossible unless some λ_σ vanishes. ■

DEFINITION 2.6. Two computations $(u_\rho \otimes v_\rho \otimes w_\rho)_{\rho \leq r}$ and $(u'_\sigma \otimes v'_\sigma \otimes w'_\sigma)_{\sigma \leq s}$ are equivalent iff $r = s$ and there is a permutation π of $\{1, \dots, r\}$ such that

$$\forall \rho \quad u_\rho \otimes v_\rho \otimes w_\rho = u'_{\pi\rho} \otimes v'_{\pi\rho} \otimes w'_{\pi\rho}.$$

THEOREM 2.7. Let $\dim U = m, \dim V = n, \dim W = q, r \leq (m-1)(n-1)$, and let t be small for $(m, n, q-1)$. Then almost all $t \in X_r \subset U \otimes V \otimes W$ have

rank r and (up to equivalence) exactly one optimal computation.

Proof.

(1) Let

$$E = \{z \in S^r : \varphi_r^{-1}\varphi_r(z) = \text{equivalence class of } z\}.$$

Clearly E is constructible. It suffices to show that E is dense in S^r . For then $\varphi_r(E)$ is constructible and dense in X_r , so almost all $t \in X_r$ have rank r (by Proposition 2.1) and lie in $\varphi_r(E)$, and hence have only one equivalence class of computations of length r .

(2) We go back to the situation described in (2.11):

$$\begin{array}{ccc} S^r & \xrightarrow{\varphi_r} & X_r \\ \downarrow \beta^r & & \downarrow \beta \\ \tilde{S}^r & \xrightarrow{\tilde{\varphi}_r} & \tilde{X}_r \end{array}$$

W.l.o.g. we may assume that $W = \tilde{W} \oplus kg$ and that α is the projection. By Proposition 2.4 applied to U, V, \tilde{W} and by Lemma 2.5, there is an open dense $\tilde{F} \subset \tilde{S}^r$ such that for any $\tilde{z} = (u_\rho \otimes v_\rho \otimes \tilde{w}_\rho)_{\rho \leq r} \in \tilde{F}$ we have

$$\tilde{\varphi}_r^{-1}\tilde{\varphi}_r(\tilde{z}) \text{ is finite,} \quad (2.14)$$

if $u \in U$, $v \in V$, and $u \otimes v \in ku_1 \otimes v_1 + \cdots + ku_r \otimes v_r$ then

$$u \otimes v \in ku_\rho \otimes v_\rho \text{ for some } \rho, \quad (2.15)$$

$$u_1 \otimes v_1, \dots, u_r \otimes v_r \text{ are linearly independent.} \quad (2.16)$$

It suffices to show that $E \cap (\beta^r)^{-1}\{\tilde{z}\}$ is dense in $(\beta^r)^{-1}\{\tilde{z}\}$ for any $\tilde{z} \in \tilde{F}$. [For then $E \cap (\beta^r)^{-1}\tilde{F}$ is dense in $(\beta^r)^{-1}\tilde{F}$ and this is open dense in S^r .]

(3) Choose $\tilde{z} = (u_\rho \otimes v_\rho \otimes \tilde{w}_\rho)_{\rho \leq r} \in \tilde{F}$. If

$$z = (u_\rho \otimes v_\rho \otimes w_\rho)_{\rho \leq r} \in (\beta^r)^{-1}\{\tilde{z}\} \setminus E,$$

then there is

$$z' = (u'_\rho \otimes v'_\rho \otimes w'_\rho)_{\rho \leq r} \in \varphi_r^{-1}\varphi_r(z)$$

not equivalent to z . We claim that

$$\sum_{\rho=1}^r ku_{\rho} \otimes v_{\rho} \not\subset \sum_{\rho=1}^r ku'_{\rho} \otimes v'_{\rho}. \quad (2.17)$$

Assume otherwise. Then because of (2.16) we have equality, and because of (2.15) we may assume (replacing z' by an equivalent computation) that

$$u'_{\rho} \otimes v'_{\rho} \in ku_{\rho} \otimes v_{\rho}$$

for all ρ . Rescaling within each $u'_{\rho} \otimes v'_{\rho} \otimes w'_{\rho}$ gives

$$u'_{\rho} \otimes v'_{\rho} = u_{\rho} \otimes v_{\rho}.$$

But then

$$\begin{aligned} \sum_{\rho} u_{\rho} \otimes v_{\rho} \otimes w_{\rho} &= \varphi_r(z) = \varphi_r(z') \\ &= \sum_{\rho} u'_{\rho} \otimes v'_{\rho} \otimes w'_{\rho} = \sum_{\rho} u_{\rho} \otimes v_{\rho} \otimes w'_{\rho}, \end{aligned}$$

so by the linear independence of the $u_{\rho} \otimes v_{\rho}$ we have $w_{\rho} = w'_{\rho}$ for all ρ , and thus $z = z'$, a contradiction. Therefore (2.17) holds. Now since $z \in (\beta')^{-1}\{\bar{z}\}$, we have

$$z = (u_{\rho} \otimes v_{\rho} \otimes w_{\rho})_{\rho \leq r} = (u_{\rho} \otimes v_{\rho} \otimes (\tilde{w}_{\rho} \oplus \lambda_{\rho} g))_{\rho \leq r}$$

with suitable $\lambda_{\rho} \in k$. Similarly

$$z' = (u'_{\rho} \otimes v'_{\rho} \otimes (\tilde{w}'_{\rho} \oplus \mu_{\rho} g))_{\rho \leq r}.$$

Writing out $\varphi_r(z) = \varphi_r(z')$, we get

$$\sum_{\rho} \lambda_{\rho} u_{\rho} \otimes v_{\rho} = \sum_{\rho} \mu_{\rho} u'_{\rho} \otimes v'_{\rho} \quad (2.18)$$

and

$$\sum_{\rho} u_{\rho} \otimes v_{\rho} \otimes \tilde{w}_{\rho} = \sum_{\rho} u'_{\rho} \otimes v'_{\rho} \otimes \tilde{w}'_{\rho},$$

or equivalently

$$(u'_\rho \otimes v'_\rho \otimes \tilde{w}'_\rho)_{\rho \leq r} \in \tilde{\varphi}_r \tilde{\varphi}_r^{-1}(\tilde{z}). \quad (2.19)$$

Combining (2.17), (2.18), and (2.19), we conclude

$$(\beta^r)^{-1}\{\tilde{z}\} \setminus E \subset \left\{ (u_\rho \otimes v_\rho \otimes (\tilde{w}_\rho \oplus \lambda_\rho g))_{\rho \leq r} : \right.$$

$\lambda_\rho \in k$ for $\rho \leq r$, and there is

$$(u'_\rho \otimes v'_\rho \otimes \tilde{w}'_\rho)_{\rho \leq r} \in \tilde{\varphi}_r^{-1} \tilde{\varphi}_r\{\tilde{z}\}$$

such that $\sum_{\rho} k u_\rho \otimes v_\rho \not\subset \sum_{\rho} k u'_\rho \otimes v'_\rho$

but $\sum_{\rho} \lambda_\rho u_\rho \otimes v_\rho \in \sum_{\rho} k u'_\rho \otimes v'_\rho$.

Now since $u_\rho, v_\rho, \tilde{w}_\rho$ are fixed and $\tilde{\varphi}_r^{-1} \tilde{\varphi}_r\{\tilde{z}\}$ is finite by (2.14), the right side of this inclusion is a proper closed subset of the irreducible variety $(\beta^r)^{-1}\{\tilde{z}\}$. Therefore $E \cap (\beta^r)^{-1}\{\tilde{z}\}$ is dense in $(\beta^r)^{-1}\{\tilde{z}\}$. ■

3. PERFECT SHAPE

We will use a Jacobian criterion to recognize certain r as being small (large) for certain shapes (m, n, q) . To this end we first compute the image of the differential of the computation map φ_r .

LEMMA 3.1. *Let $x = (u_\rho \otimes v_\rho \otimes w_\rho)_{\rho \leq r} \in S^r$ be such that no $u_\rho \otimes v_\rho \otimes w_\rho$ vanishes. Then*

$$\text{im } d_x \varphi_r = \sum_{\rho=1}^r (U \otimes v_\rho \otimes w_\rho + u_\rho \otimes V \otimes w_\rho + u_\rho \otimes v_\rho \otimes W).$$

(Here $U \otimes v_\rho \otimes w_\rho$ denotes the linear subspace of $U \otimes V \otimes W$ of all tensors of the form $u \otimes v_\rho \otimes w_\rho$, $u \in U$.)

Proof. The differential of the tensor map

$$\begin{aligned}\tau: U \times V \times W &\rightarrow S, \\ (u, v, w) &\rightarrow u \otimes v \otimes w\end{aligned}$$

at any point (u, v, w) such that $u \otimes v \otimes w \neq 0$ is surjective, since locally at such points τ has a section (= right inverse). A similar statement then holds for the differential of the map

$$\tau': (U \times V \times W)^r \rightarrow S^r.$$

Let

$$\begin{aligned}\psi_r: (U \times V \times W)^r &\rightarrow U \otimes V \otimes W \\ (u_\rho, v_\rho, w_\rho)_{\rho \leq r} &\mapsto \sum_{\rho=1}^r u_\rho \otimes v_\rho \otimes w_\rho\end{aligned}$$

be the composition of τ' with φ_r . By the above it suffices to show

$$\text{im } d_x \psi_r = \sum_{\rho=1}^r (U \otimes v_\rho \otimes w_\rho + u_\rho \otimes V \otimes w_\rho + u_\rho \otimes v_\rho \otimes W),$$

where $x = (u_\rho, v_\rho, w_\rho)_{\rho \leq r}$ and no u_ρ, v_ρ, w_ρ vanishes. But ψ_r is a polynomial map between vector spaces, and its differential is easily computed via the Jacobian matrix:

$$(d_x \psi_r)(x_\rho, y_\rho, z_\rho)_{\rho \leq r} = \sum_{\rho=1}^r (x_\rho \otimes v_\rho \otimes w_\rho + u_\rho \otimes y_\rho \otimes w_\rho + u_\rho \otimes v_\rho \otimes z_\rho).$$

This proves the lemma. ■

Since $\dim S^r = r(m+n+q-2)$ and S is smooth except at the origin, it is clear that

$$\begin{aligned}\dim \sum_{\rho=1}^r (U \otimes v_\rho \otimes w_\rho + u_\rho \otimes V \otimes w_\rho + u_\rho \otimes v_\rho \otimes W) \\ \leq \min \{r(m+n+q-2), mnq\}.\end{aligned}\tag{3.1}$$

The next proposition relates equality here to equality in (2.9).

PROPOSITION 3.2. *If there is $(u_\rho \otimes v_\rho \otimes w_\rho)_{\rho \leq r} \in S^r$ such that*

$$\dim \left(\sum_{\rho=1}^r U \otimes v_\rho \otimes w_\rho + u_\rho \otimes V \otimes w_\rho + u_\rho \otimes v_\rho \otimes W \right) = r(m+n+q-2), \quad (3.2)$$

then r is small. If there is $(u_\rho \otimes v_\rho \otimes w_\rho)_{\rho \leq r} \in S^r$ such that

$$\dim \sum_{\rho=1}^r (U \otimes v_\rho \otimes w_\rho + u_\rho \otimes V \otimes w_\rho + u_\rho \otimes v_\rho \otimes W) = mnq, \quad (3.3)$$

then r is large.

Proof. Let $c \in \mathbb{N}$ and

$$\dim \sum_{\rho=1}^r (U \otimes v_\rho \otimes w_\rho + u_\rho \otimes V \otimes w_\rho + u_\rho \otimes v_\rho \otimes W) \geq c \quad (3.4)$$

for some $x = (u_\rho \otimes v_\rho \otimes w_\rho)_{\rho \leq r} \in S^r$. Using the determinantal criterion for linear independence, we see that this inequality holds for almost all $x \in S^r$. Thus we may assume that $x \in S^r$ and $\varphi_r(x) \in X_r$ are nonsingular points. (Evidently x is nonsingular iff no $u_\rho \otimes v_\rho \otimes w_\rho$ vanishes.) But then

$$\begin{aligned} \dim X_r &= \dim (\text{tangent space of } X_r \text{ at } \varphi_r(x)) \\ &\geq \dim \text{im} (d_x \varphi_r) \geq c \end{aligned} \quad (3.5)$$

by Lemma 3.1. Now take $c = r(m+n+q-2)$ and $c = mnq$ respectively and use (2.9). ■

If $\text{char } k = 0$ then the conditions of the previous proposition are necessary and sufficient for r to be small or large, respectively. These conditions are easier to verify than those of Definition 2.2, since they concern the dimension of a linear subspace of $U \otimes V \otimes W$ instead of a subvariety. To compute the dimension of this linear space in certain cases, we use a splitting technique, exemplified in the next two lemmas and the subsequent theorem.

LEMMA 3.3. Let $\dim U = m$, $\dim V = n$, $\dim W = 2$, $2 \leq s \leq m \leq n$, and $r \geq s$. Then

$$\dim \left(\sum_{\sigma=1}^s U \otimes v_{\sigma} \otimes w_{\sigma} + u_{\sigma} \otimes V \otimes w_{\sigma} + u_{\sigma} \otimes v_{\sigma} \otimes W + \sum_{\rho=s+1}^r u_{\rho} \otimes v_{\rho} \otimes W \right) \\ = \min \{ (m+n)s + 2(r-s), 2mn \}$$

for almost all $u_1, \dots, u_r \in U$, $v_1, \dots, v_r \in V$, $w_1, \dots, w_s \in W$.

The lemma states that the dimension of the linear subspace referred to (which appears as subproblem in the splitting process) is as large as it can possibly be. Its proof is similar to, but much easier than, the proof of the next lemma and is therefore omitted.

LEMMA 3.4. Let $\dim U = m$, $\dim V = n$, $\dim W = 3$, $3 \leq s \leq m \leq n$, and $r \geq s$. Then

$$\dim \left(\sum_{\sigma=1}^s U \otimes v_{\sigma} \otimes w_{\sigma} + u_{\sigma} \otimes V \otimes w_{\sigma} + u_{\sigma} \otimes v_{\sigma} \otimes W + \sum_{\rho=s+1}^r u_{\rho} \otimes v_{\rho} \otimes W \right) \\ = \min \{ (m+n+1)s + 3(r-s), 3mn \} \quad (3.6)$$

for almost all $u_1, \dots, u_r \in U$, $v_1, \dots, v_r \in V$, $w_1, \dots, w_s \in W$.

Proof. By the determinantal criterion for linear independence it suffices to show the existence of $u_1, \dots, u_r \in U$, $v_1, \dots, v_r \in V$, and $w_1, \dots, w_s \in W$ such that (3.6) holds.

If the u_i, v_i, w_i are already chosen, we use the abbreviation

$$Y_r = \sum_{\sigma \leq s} (U \otimes v_{\sigma} \otimes w_{\sigma} + u_{\sigma} \otimes V \otimes w_{\sigma} + u_{\sigma} \otimes v_{\sigma} \otimes W) + \sum_{s < \rho \leq r} u_{\rho} \otimes v_{\rho} \otimes W.$$

Let r_0 be such that

$$(m+n+1)s + 3(r_0 - 1 - s) < 3mn \leq (m+n+1)s + 3(r_0 - s).$$

Note that $r_0 > s$. Since

$$\dim Y_s \leq (m+n+1)s$$

and

$$\dim(u_\rho \otimes v_\rho \otimes W) \leq 3,$$

it suffices to prove the lemma for $r = r_0 - 1$ and $r = r_0$. [Or just for $r = r_0$ in the event $3mn = (m + n + 1)s + 3(r_0 - s)$.]

Let e_1, \dots, e_m and f_1, \dots, f_n be bases of U and V respectively. For $1 \leq i \leq s$ we choose once and for all

$$u_i = e_i, \quad v_i = f_i$$

and $w_i \in W$ such that any three of the w_i are linearly independent and such that

$$(kw_{i_1} + kw_{j_1}) \cap (kw_{i_2} + kw_{j_2}) \cap (kw_{i_3} + kw_{j_3}) = \{0\} \quad (3.7)$$

whenever $i_1, j_1, i_2, j_2, i_3, j_3$ are all distinct. (Such a choice is possible, since each (3.7) is an open condition when the w_i are linearly independent in pairs.) Then we have

$$\begin{aligned} Y_s = & \bigoplus_{i \leq s} e_i \otimes f_i \otimes W \oplus \bigoplus_{\substack{i, j \leq s \\ i \neq j}} e_i \otimes f_j \otimes (kw_i + kw_j) \\ & \oplus \bigoplus_{\substack{i \leq s \\ j > s}} e_i \otimes f_j \otimes kw_i \oplus \bigoplus_{\substack{i > s \\ j \leq s}} e_i \otimes f_j \otimes kw_j \end{aligned}$$

(where we tacitly always assume $i \leq m, j \leq n$); in particular

$$\dim Y_s = (m + n + 1)s.$$

If we put

$$W_{ij} = \begin{cases} kw_i + kw_j & \text{if } i, j \leq s, i \neq j, \\ kw_i & \text{if } i \leq s, j > s, \\ kw_j & \text{if } i > s, j \leq s, \\ 0 & \text{if } i > s, j > s \end{cases}$$

and $\Delta = \{(i, i) : 1 \leq i \leq s\}$, then we can write

$$Y_s = \bigoplus_{(i, j) \in \Delta} e_i \otimes f_j \otimes W \oplus \bigoplus_{(i, j) \notin \Delta} e_i \otimes f_j \otimes W_{ij}.$$

For $\Delta \subset T \subset [1, m] \times [1, n]$ (where $[1, m] = \{1, \dots, m\}$) we define

$$Z_T = \bigoplus_{(i, j) \in T} e_i \otimes f_j \otimes W \oplus \bigoplus_{(i, j) \notin T} e_i \otimes f_j \otimes W_{ij}.$$

Then of course $Y_s = Z_\Delta$. Let us call a set $T \subset [1, m] \times [1, n]$ *reachable*, if there are $t \geq s$, $u_{s+1}, \dots, u_t \in U$, and $v_{s+1}, \dots, v_t \in V$ such that

$$Y_t = Z_T, \quad \dim Z_T = (m + n + 1)s + 3(t - s).$$

If T is reachable, then the lemma is correct for $r = t$. Δ is reachable. We try to construct reachable sets from reachable ones. To do this we use the following terminology for points $(i, j) \in [1, m] \times [1, n]$:

- (i, j) and (j, i) lie symmetrically.
- (i, j) is red iff $i, j \leq s$, $i \neq j$.
- (i, j) is yellow iff $i \leq s$, $j > s$ or $i > s$, $j \leq s$.
- (i, j) is blue iff $i > s$, $j > s$.

A cross is a set of the form

$$K_i = \{(i, j) : j \leq n\} \cup \{(j, i) : j \leq m\}$$

for some $i \leq s$.

We will prove the following assertion:

(@) Let T be reachable, $A \subset [1, m]$, $B \subset [1, n]$, and let $(A \times B) \setminus T$ consist of either

- (1) three red points, which do not all lie on the same cross and no two of which lie symmetrically, or
- (2) one red and one yellow point, which do not both lie on the same cross, or
- (3) three yellow points, no two of which lie on the same cross, or
- (4) one blue point.

Then $T \cup (A \times B)$ is reachable.

To prove this let $t \geq s$, $u_{s+1}, \dots, u_t \in U$, $v_{s+1}, \dots, v_t \in V$ such that

$$Y_t = Z_T, \quad \dim Z_T = (m + n + 1)s + 3(t - s).$$

We treat the cases (1)–(4) separately.

Case (1): Let

$$(A \times B) \setminus T = \{(i_1, j_1), (i_2, j_2), (i_3, j_3)\}.$$

We take

$$u_{t+1} = \sum_{i \in A} e_i, \quad v_{t+1} = \sum_{j \in B} f_j.$$

Then we have

$$\begin{aligned} Y_{t+1} &= Z_T + u_{t+1} \otimes v_{t+1} \otimes W \\ &= \bigoplus_{(i,j) \in T} e_i \otimes f_j \otimes W \oplus \bigoplus_{(i,j) \notin T \cup (A \times B)} e_i \otimes f_j \otimes W_{ij} \oplus P, \end{aligned}$$

where

$$P = \bigoplus_{\nu=1}^3 e_{i_\nu} \otimes f_{j_\nu} \otimes W_{i_\nu, j_\nu} + \left(\sum_{\nu=1}^3 e_{i_\nu} \otimes f_{i_\nu} \right) \otimes W.$$

It suffices to show that

$$\left(\bigoplus_{\nu=1}^3 e_{i_\nu} \otimes f_{j_\nu} \otimes W_{i_\nu, j_\nu} \right) \cap \left(\sum_{\nu=1}^3 e_{i_\nu} \otimes f_{j_\nu} \right) \otimes W = \{0\}. \quad (3.8)$$

[For then $\dim P = 9$, thus

$$P = \bigoplus_{\nu=1}^3 e_{i_\nu} \otimes f_{j_\nu} \otimes W,$$

and thus

$$Y_{t+1} = Z_{T \cup (A \times B)}$$

and

$$\dim Z_{T \cup (A \times B)} = \dim Z_T + 3 = (m + n + 1)s + 3(t + 1 - s).]$$

(3.8) is equivalent to

$$W_{i_1j_1} \cap W_{i_2j_2} \cap W_{i_3j_3} = \{0\}. \quad (3.9)$$

First we show $W_{i_\mu j_\mu} \neq W_{i_\nu j_\nu}$ for $\mu \neq \nu$. By way of contradiction assume e.g. $W_{i_1j_1} = W_{i_2j_2}$, i.e.

$$kw_{i_1} + kw_{j_1} = kw_{i_2} + kw_{j_2}.$$

Since any three of the w_i are linearly independent, this implies

$$\{i_1, j_1\} = \{i_2, j_2\}.$$

So either $(i_1, j_1) = (i_2, j_2)$ or (i_1, j_1) and (i_2, j_2) lie symmetrically, which is impossible. Next we show (3.9). Assuming (3.9) to be wrong, we conclude from (3.7) that the sets $\{i_\mu, j_\mu\}$ are not pairwise disjoint. [$i_\mu \neq j_\mu$ because (i_μ, j_μ) is red.] Let e.g. $i_1 \in \{i_2, j_2\}$. Then

$$kw_{i_1} \subset W_{i_1j_1} \cap W_{i_2j_2},$$

together with $W_{i_1j_1} \neq W_{i_2j_2}$. Therefore

$$kw_{i_1} = W_{i_1j_1} \cap W_{i_2j_2}.$$

Since (3.9) is assumed to be incorrect, we get

$$kw_{i_1} = W_{i_1j_1} \cap W_{i_2j_2} \cap W_{i_3j_3},$$

thus

$$w_{i_1} \in kw_{i_3} + kw_{j_3},$$

and thus

$$i_1 \in \{i_3, j_3\}.$$

But then all three points (i_ν, j_ν) lie on the cross K_{i_1} , a contradiction.

Case (2): Similar and simpler.

Case (3): Let

$$(A \times B) \setminus T = \{(i_1, j_1), (i_2, j_2), (i_3, j_3)\}.$$

We take

$$\begin{aligned} u_{t+1} &= \sum_{i \in A} e_i, & v_{t+1} &= \sum_{j \in B} f_j, \\ u_{t+2} &= \sum_{i \in A} \alpha_i e_i, & v_{t+2} &= \sum_{j \in B} \beta_j f_j, \end{aligned}$$

where $\alpha_i, \beta_j \in k$, $\alpha_{i_\mu} \beta_{j_\mu} \neq \alpha_{i_\nu} \beta_{j_\nu}$ for $\mu \neq \nu$. Then

$$\begin{aligned} Y_{t+2} &= Z_T + u_{t+1} \otimes v_{t+1} \otimes W + u_{t+2} \otimes v_{t+2} \otimes W \\ &= \bigoplus_{(i,j) \in T} e_i \otimes f_j \otimes W \oplus \bigoplus_{(i,j) \notin T \cup (A \times B)} e_i \otimes f_j \otimes W_{ij} \oplus Q, \end{aligned}$$

where

$$Q = \bigoplus_{\nu=1}^3 e_{i_\nu} \otimes f_{j_\nu} \otimes W_{i_\nu, j_\nu} + \left[\left(\sum_{\nu=1}^3 e_{i_\nu} \otimes f_{j_\nu} \right) \otimes W \oplus \left(\sum_{\nu=1}^3 \alpha_{i_\nu} \beta_{j_\nu} e_{i_\nu} \otimes f_{j_\nu} \right) \otimes W \right].$$

Again it suffices to show that this sum is direct, in other words that

$$\begin{aligned} &\left(\bigoplus_{\nu=1}^3 e_{i_\nu} \otimes f_{j_\nu} \otimes W_{i_\nu, j_\nu} \right) \\ &\cap \left[\left(\sum_{\nu} e_{i_\nu} \otimes f_{j_\nu} \right) \otimes W \oplus \left(\sum_{\nu} \alpha_{i_\nu} \beta_{j_\nu} e_{i_\nu} \otimes f_{j_\nu} \right) \otimes W \right] = \{0\}. \end{aligned}$$

Now if $\sum_{\nu} e_{i_\nu} \otimes f_{j_\nu} \otimes x_{\nu}$ lies in the above intersection, then on the one hand

$$\dim(kx_1 + kx_2 + kx_3) = \#\{\nu : x_{\nu} \neq 0\},$$

on the other hand

$$\dim(kx_1 + kx_2 + kx_3) \leq \#\{\nu : x_{\nu} \neq 0\} - 1$$

unless $x_1 = x_2 = x_3 = 0$. Therefore $x_1 = x_2 = x_3 = 0$.

Case (4): Trivial.

This proves the assertion (Q).

We now prove the lemma for $s \equiv 0 \pmod{3}$ by showing that in this case $[1, m] \times [1, n]$ is reachable. (The reader is advised to illustrate the following considerations by a picture of the "matrix" $[1, m] \times [1, n]$, say for $m = 10, n = 12, s = 9$.)

Starting with Δ , we first adjoin [according to (1)] the sets

$$A_i \times B_{ai} = \{i \bmod s, i + 1 \bmod s\} \times \{i + 3\alpha - 2 \bmod s, i + 3\alpha \bmod s\} \\ \left(1 \leq \alpha \leq \frac{s}{3} - 1, \quad 1 \leq i \leq s\right)$$

in lexicographical order with respect to (α, i) . (Here and in the sequel $j \bmod s$ means the residue of j modulo s in the system of residues $\{1, \dots, s\}$.) By induction on α one sees that

$$T_\alpha = \{(i \bmod s, j \bmod s) : 0 \leq j - i \leq 3\alpha\}$$

is reachable. (Namely, by using those $A_i \times B_{ai}$ with $1 \leq \omega \leq \alpha, 1 \leq i \leq s$. Note that for a given α the $A_i \times B_{ai}$ are pairwise disjoint and each contains exactly one point in $T_{\alpha-1}$.) In particular

$$T = \{(i \bmod s, j \bmod s) : 0 \leq j - i \leq s - 3\}$$

is reachable. Next we adjoin the sets

$$\{(3\beta - 1 \bmod s, 3\beta \bmod s) \times (3\beta - 2 \bmod s, 3\beta - 1 \bmod s) \quad (1 \leq \beta \leq s/3)$$

and

$$\{3\gamma + 1 \bmod s, 3\gamma + 2 \bmod s\} \times \{3\gamma - 1 \bmod s, 3\gamma \bmod s\} \quad (1 \leq \gamma \leq s/3).$$

(Again these sets are pairwise disjoint and each contains exactly one element of T .) This shows that $[1, s] \times [1, s]$ is reachable. Finally we adjoin the sets

$$\{3\delta - 2, 3\delta - 1, 3\delta\} \times \{j\} \quad (1 \leq \delta \leq s/3, \quad j > s)$$

and

$$\{i\} \times \{3\varepsilon - 2, 3\varepsilon - 1, 3\varepsilon\} \quad (i > s, \quad 1 \leq \varepsilon \leq s/3)$$

[according to (3)] and the sets

$$\{i\} \times \{j\} \quad (i > s, \quad j > s)$$

according to (4). Then $[1, m] \times [1, n]$ is reached and the lemma is proved for $r = r_0$. This settles the case $s \equiv 0 \pmod{3}$.

Now we consider the case $s \equiv 1 \pmod{3}$. (Make an illustration for $m = 14$, $n = 18$, $s = 10$.) To Δ we adjoin the sets

$$\begin{aligned} & \{i \bmod s, i + 1 \bmod s\} \times \{i + 3\alpha - 2 \bmod s, i + 3\alpha \bmod s\} \\ & \left(1 \leq \alpha \leq \frac{s-1}{3}, \quad i \leq i \leq s \right) \end{aligned}$$

in lexicographical order with respect to (α, i) . Thereby we have reached $[1, s] \times [1, s]$. Let $j_h = j - s \bmod 3$ in the system of residues $\{1, 2, 3\}$. We adjoin the sets

$$\begin{aligned} & ((1, 2, 3, 4) \setminus \{j_h\}) \times \{j\}, \quad (3\beta - 1, 3\beta, 3\beta + 1) \times \{j\} \\ & \left(j > s, \quad 2 \leq \beta \leq \frac{s-1}{3} \right), \\ & \{1, 2, 3\} \times \{s + 3\gamma - 2, s + 3\gamma - 1, s + 3\gamma\} \quad (\gamma \geq 1, \quad s + 3\gamma \leq n). \end{aligned}$$

This takes care of the right upper $s \times (n - s)$ rectangle of $[1, m] \times [1, n]$ except for at most two (yellow) points on the first two rows. Similarly let $i_1 = i - s \bmod 3 \in \{1, 2, 3\}$. We adjoin

$$\begin{aligned} & \{i\} \times (\{s - 3, s - 2, s - 1, s\} \setminus \{s - i_1 + 1\}), \quad \{i\} \times \{3\delta - 2, 3\delta - 1, 3\delta\} \\ & \left(i > s, \quad 1 \leq \delta \leq \frac{s-1}{3} - 1 \right), \\ & \{s + 3\epsilon - 2, s + 3\epsilon - 1, s + 3\epsilon\} \times \{s - 2, s - 1, s\} \quad (\epsilon \geq 1, \quad s + 3\epsilon \leq m), \end{aligned}$$

and finally

$$\{i\} \times \{j\} \quad (i, j > s).$$

Then everything is reached except for at most four yellow points, no two of which lie on a cross. An easy argument now proves the lemma for $s \equiv 1 \pmod{3}$.

We turn to the case $s \equiv 2 \pmod{3}$ (look at $m = 11$, $n = 14$, $s = 8$), where we first adjoin the sets

$$A_i \times B_{\alpha i} \quad \left(1 \leq \alpha \leq \frac{s-2}{3}, \quad 1 \leq i \leq s \right)$$

in lexicographical order for (α, i) , and the sets

$$\{3\beta, 3\beta+1, 3\beta+2\} \times \{3\beta-1, 3\beta, 3\beta+1\} \quad \left(1 \leq \beta \leq \frac{s-2}{3} \right).$$

This reaches $[1, s] \times [1, s]$ except for the points $(2, 1)$ and $(1, s)$. If $n = s$ (and therefore also $m = s$), the proof of the lemma is easily completed. So assume $n > s$. We adjoin successively

$$\{3\gamma-2, 3\gamma-1, 3\gamma\} \times \{j\} \quad \left(1 \leq \gamma \leq \frac{s-2}{3}, \quad j-s > 0 \text{ odd} \right),$$

$$\{3\delta, 3\delta+1, 3\delta+2\} \times \{j\} \quad \left(1 \leq \delta \leq \frac{s-2}{3}, \quad j-s > 0 \text{ even} \right),$$

$$\{1, s-1\} \times \{s, s+1\}, \quad \{2, s\} \times \{1, s+1\}$$

[here we use (2) for the only time],

$$\{2, s-1, s\} \times \{j, j+1\} \quad (s < j < n, \quad j-s \equiv 2 \pmod{6}),$$

$$\{1, s-1, s\} \times \{j, j+1\} \quad (s < j < n, \quad j-s \equiv 3 \pmod{6}),$$

$$\{1, 2, s-1\} \times \{j, j+1\} \quad (s < j < n, \quad j-s \equiv 5 \pmod{6}),$$

$$\{1, 2, s\} \times \{j, j+1\} \quad (s < j < n, \quad j-s \equiv 0 \pmod{6}).$$

This covers the right upper $s \times (n-s)$ rectangle except for at most two points on rows 1 and 2 or rows $s-1$ and s . The left lower $(m-s) \times s$ rectangle is treated similarly, and the right lower $(m-s) \times (n-s)$ rectangle in the obvious way. One only has to take care that no two of the at most four yellow points that remain uncovered are on a cross. Then the lemma follows as before. ■

THEOREM 3.5. *Let $2 \leq m \leq n \leq m + q - 2$, q be even, and $q - 1 \leq (m - 1)(n - 1)$. Then*

$$\begin{aligned} \frac{q}{2} \left\lfloor \frac{2mn}{m+n+q-2} \right\rfloor & \text{ is small,} \\ \frac{q}{2} \left\lceil \frac{2mn}{m+n+q-2} \right\rceil & \text{ is large.} \end{aligned}$$

Proof. First assertion: Let

$$s = \left\lfloor \frac{2mn}{m+n+q-2} \right\rfloor, \quad r = \frac{q}{2}s,$$

and write

$$W = W_1 \oplus \cdots \oplus W_{q/2}, \quad \dim W_i = 2 \quad \text{for all } i.$$

Given $u_1, \dots, u_r \in U$, $v_1, \dots, v_r \in V$, $w_1, \dots, w_r \in W$, we will use the notation

$$u_{\kappa, \sigma} = u_{(\kappa-1)s+\sigma}, \quad v_{\kappa, \sigma} = v_{(\kappa-1)s+\sigma}, \quad w_{\kappa, \sigma} = w_{(\kappa-1)s+\sigma}$$

for $\kappa = 1, \dots, q/2$, $\sigma = 1, \dots, s$. Suppose

$$w_{\kappa, \sigma} \in W_{\kappa} \quad \text{for all } \kappa, \sigma.$$

Then

$$\begin{aligned} & \sum_{\rho=1}^r (U \otimes v_{\rho} \otimes w_{\rho} + u_{\rho} \otimes V \otimes w_{\rho} + u_{\rho} \otimes v_{\rho} \otimes W) \\ &= \bigoplus_{\kappa=1}^{q/2} \left(\sum_{\sigma=1}^s (U \otimes v_{\kappa, \sigma} \otimes w_{\kappa, \sigma} + u_{\kappa, \sigma} \otimes V \otimes w_{\kappa, \sigma} + u_{\kappa, \sigma} \otimes v_{\kappa, \sigma} \otimes W_{\kappa}) \right. \\ & \quad \left. + \sum_{\substack{\lambda \neq \kappa \\ 1 \leq \sigma \leq s}} u_{\lambda, \sigma} \otimes v_{\lambda, \sigma} \otimes W_{\kappa} \right) \\ &= \bigoplus_{\kappa=1}^{q/2} I_{\kappa} \quad (\text{say}). \end{aligned}$$

By Lemma 3.3 for almost all choices $u_{\kappa, \sigma} \in U$, $v_{\kappa, \sigma} \in V$, $w_{\kappa, \sigma} \in W_{\kappa}$ ($\kappa = 1, \dots$,

$q/2, \sigma = 1, \dots, s)$ we have

$$\begin{aligned}\dim I_{\kappa} &= \min\{(m+n)s + 2(r-s), 2mn\} \\ &= (m+n+q-2)s \quad \text{for all } \kappa;\end{aligned}$$

hence

$$\dim \bigoplus_{\kappa=1}^{q/2} I_{\kappa} = (m+n+q-2)r.$$

Proposition 3.2 now says that r is small.

Second assertion: Similarly, with $s = \lceil 2mn/(m+n+q-2) \rceil$. ■

COROLLARY 3.6. *Let $2 \leq m \leq n \leq m+q-2$, $q-1 \leq (m-1)(n-1)$, and q be even. Then*

$$\frac{mnq}{m+n+q-2} \leq \underline{R}(m, n, q) < \frac{mnq}{m+n+q-2} + \frac{q}{2}.$$

Proof. The left inequality follows from (2.8), the right inequality from

$$\begin{aligned}\underline{R}(m, n, q) &\leq \frac{q}{2} \left\lceil \frac{2mn}{m+n+q-2} \right\rceil \\ &< \frac{q}{2} \left(\frac{2mn}{m+n+q-2} + 1 \right).\end{aligned}\quad \blacksquare$$

COROLLARY 3.7. *Let $3 \leq m \leq n \leq q$, $q-1 \leq (m-1)(n-1)$, and q be odd. If*

$$r \leq \frac{mnq}{m+n+q-2} - q,$$

then almost all tensors $t \in X_r$ have, up to equivalence, exactly one optimal computation.

Proof. We apply Theorem 2.7. The assumption $r \leq (m-1)(n-1)$ follows from

$$\begin{aligned}r(m+n+q-2) &\leq mnq - q(m+n+q-2) \\ &\leq (mn - m - n)q \leq (m-1)(n-1)(m+n+q-2).\end{aligned}$$

The assumption that r is small for $(m, n, q-1)$ follows from Theorem 3.5, as applied to $(m, n, q-1)$. ■

DEFINITION 3.8. A shape (m, n, q) is good iff

$$\forall r \quad \dim X_r = \min\{r(m+n+q-2), mnq\}.$$

(m, n, q) is perfect iff it is good and $mnq/(m+n+q-2)$ is an integer.

Obviously, (m, n, q) is good iff any r is either small or large for (m, n, q) . By Proposition 2.3 this is the case iff $\lfloor mnq/(m+n+q-2) \rfloor$ is small and $\lceil mnq/(m+n+q-2) \rceil$ is large. Thus (m, n, q) is perfect iff $mnq/(m+n+q-2)$ is an integer large for (m, n, q) . In this case there is a d such that almost all $t \in U \otimes V \otimes W$ have rank $r = mnq/(m+n+q-2)$ and exactly d optimal computations.

Let us call a shape (m, n, q) balanced iff

$$\begin{aligned} m-1 &\leq (n-1)(q-1), \\ n-1 &\leq (m-1)(q-1), \\ q-1 &\leq (m-1)(n-1); \end{aligned} \tag{3.10}$$

otherwise unbalanced. Assuming w.l.o.g. $2 \leq m \leq n \leq q$, this is equivalent to

$$q-1 \leq (m-1)(n-1), \tag{3.11}$$

and also to

$$\frac{mnq}{m+n+q-2} \geq q. \tag{3.12}$$

Unbalanced shapes cannot be perfect. For otherwise let (m, n, q) be unbalanced and perfect, $m \leq n \leq q$, and put $r = mnq/(m+n+q-2)$. Then $r < \min\{q, mn\}$. Choose bases $e_1, \dots, e_m \in U$, $f_1, \dots, f_n \in V$. Since (m, n, q) is perfect, almost all tensors $t = \sum_{i,j} e_i \otimes f_j \otimes t_{ij} \in U \otimes V \otimes W$ have rank r . But then for almost all t we have

$$\min\{q, mn\} = \dim\{t_{ij}; i \leq m, j \leq n\} \leq r,$$

which is absurd.

I do not know the answer to the following interesting problem: Are all balanced shapes (m, n, q) , such that $mnq/(m+n+q-2)$ is an integer,

perfect? We shall see in the next section that not all balanced shapes are good (e.g., $(3, 3, 3)$ is not good).

THEOREM 3.9. *Let (m, n, q) be balanced, $m \leq n$. Then (m, n, q) is perfect provided that any of the following conditions is satisfied:*

$$q \text{ even and } 2n \leq m + n + q - 2 \mid 2mn. \quad (1)$$

$$3 \mid q \text{ and } 3n \leq m + n + q - 2 \mid 3mn. \quad (2)$$

$$3n \leq m + n + q - 2 \mid mn. \quad (3)$$

Proof. (1) follows from Theorem 3.5. (2) is proved in a similar way by splitting up W into $q/3$ subspaces of dimension 3 each (compare the proof of Theorem 3.5) and using Lemma 3.4 instead of Lemma 3.3. For (3) we may assume q to be odd and $q \geq 3$. Now split up W into $(q-3)/2$ subspaces of dimension 2 accommodating $s_1 = 2mn/(m+n+q-2)$ triads and one subspace of dimension 3 accommodating $s_2 = 3mn/(m+n+q-2)$ triads, and use Lemmas 3.3 and 3.4 respectively.

COROLLARY 3.10.

(1) $(n, n, n+2)$ is perfect for $n \not\equiv 2 \pmod{3}$.

(2) $(n-1, n, n)$ is perfect for $n \equiv 0 \pmod{6}$.

Actually $(n-1, n, n)$ is perfect as long as $n \equiv 0 \pmod{3}$. The proof of this requires some extra work, however. The next corollary shows that perfect shapes are relatively dense.

COROLLARY 3.11. *Let $\alpha \leq \beta \leq \gamma$ be positive integers. Then $(\alpha j, \beta j, \gamma j + 2)$ is perfect for any positive integer j such that*

$$j \equiv 0 \pmod{2(\alpha + \beta + \gamma)}.$$

4. A DETERMINANT FOR 3-SLICE TENSORS

Here we assume $\dim U = \dim V = n$, $\dim W = 3$. We choose bases (e_1, \dots, e_n) for U , (f_1, \dots, f_n) for V , (g_1, g_2, g_3) for W . If $t \in U \otimes V \otimes W$ we write

$$t = \sum_{i,j} \alpha_{ij} e_i \otimes f_j \otimes g_1 + \sum_{i,j} \beta_{ij} e_i \otimes f_j \otimes g_2 + \sum_{i,j} \gamma_{ij} e_i \otimes f_j \otimes g_3 \quad (4.1)$$

and define the $n \times n$ matrices

$$A = (\alpha_{ij}), \quad B = (\beta_{ij}), \quad C = (\gamma_{ij}).$$

A, B, C are the three "slices" of t .

THEOREM 4.1. *Let A be invertible. Then*

$$\underline{R}(t) \geq n + \frac{1}{2} \text{rank}(BA^{-1}C - CA^{-1}B).$$

Proof. We first show

$$R(t) \geq n + \frac{1}{2} \text{rank}(BA^{-1}C - CA^{-1}B). \quad (4.2)$$

W.l.o.g. we may assume that A is the unit matrix $\mathbb{1}_n$. (The tensor with slices $\mathbb{1}_n, A^{-1}B, A^{-1}C$ is isomorphic to t .) Suppose

$$R(t) \leq r,$$

i.e.

$$t = \sum_{\rho=1}^r u_{\rho} \otimes v_{\rho} \otimes w_{\rho} \quad (4.3)$$

for some $u_{\rho} \in U, v_{\rho} \in V, w_{\rho} \in W$. We then have to show

$$\text{rank}(BC - CB) \leq 2(r - n). \quad (4.4)$$

Let

$$u_{\rho} = \sum_{i=1}^n \eta_{\rho i} e_i, \quad v_{\rho} = \sum_{i=1}^n \xi_{\rho i} f_i, \quad w_{\rho} = \sum_{l=1}^3 \theta_{\rho l} g_l,$$

and introduce the $r \times n$ matrices

$$H = (\eta_{\rho i}), \quad Z = (\xi_{\rho j})$$

and the diagonal matrices

$$T_l = \begin{pmatrix} \theta_{1l} & & 0 \\ & \ddots & \\ 0 & & \theta_{rl} \end{pmatrix}, \quad l \in \{1, 2, 3\}.$$

Comparing coefficients in (4.1) and (4.3), we get

$$\begin{aligned} H^T T_1 Z &= A = \mathbf{1}_n, \\ H^T T_2 Z &= B, \\ H^T T_3 Z &= C, \end{aligned} \tag{4.5}$$

(H^T denotes the transpose of H .) Assume first that no $\theta_{\rho 1}$ vanishes. Then w.l.o.g. $T_1 = \mathbf{1}_r$, the unit matrix. (Replace $\zeta_{\rho j}$ by $\theta_{\rho 1} \zeta_{\rho j}$ and $\theta_{\rho l}$ by $\theta_{\rho 1}^{-1} \theta_{\rho l}$.) Thus

$$H^T Z = \mathbf{1}_n.$$

Augment Z to an $r \times r$ matrix \hat{Z} by attaching $r - n$ columns to the right side orthogonal to the rows of H^T . Then \hat{Z}^{-1} is obtained from H^T by attaching $r - n$ rows to the bottom. If we define

$$\hat{B} = \hat{Z}^{-1} T_2 \hat{Z}, \quad \hat{C} = \hat{Z}^{-1} T_3 \hat{Z},$$

we therefore get from (4.5)

$$\hat{B} = \begin{pmatrix} B & B_{12} \\ B_{21} & B_{22} \end{pmatrix}, \quad \hat{C} = \begin{pmatrix} C & C_{12} \\ C_{21} & C_{22} \end{pmatrix}.$$

Now

$$\hat{B}\hat{C} - \hat{C}\hat{B} = \hat{Z}^{-1} T_2 T_3 \hat{Z} - \hat{Z}^{-1} T_3 T_2 \hat{Z} = 0.$$

So

$$BC + B_{12}C_{21} - CB - C_{12}B_{21} = 0,$$

and therefore

$$\text{rank}(BC - CB) = \text{rank}(C_{12}B_{21} - B_{12}C_{21}) \leq 2(r - n),$$

since C_{12} and B_{12} have only $r - n$ columns. Thus we have shown (4.4) under the additional hypothesis that no $\theta_{\rho 1}$ vanishes. We complete the proof by induction on p , the number of ρ such that $\theta_{\rho 1} = 0$. Let $p > 0$, W.l.o.g. $\theta_{r1} = 0$.

Put

$$\tilde{t} = \sum_{\rho=1}^{r-1} u_{\rho} \otimes v_{\rho} \otimes w_{\rho}.$$

Let $\tilde{A}, \tilde{B}, \tilde{C}$ be the slices of \tilde{t} . Then $\tilde{A} = \mathbf{1}_n$. By induction hypothesis we therefore have

$$\text{rank}(\tilde{B}\tilde{C} - \tilde{C}\tilde{B}) \leq 2(r-1-n).$$

Moreover

$$B = \tilde{B} + \theta_{r_2} D, \quad C = \tilde{C} + \theta_{r_3} D,$$

where $D = (\eta_{ri} \zeta_{ri})$ is a matrix of rank 1. Thus

$$\text{rank}(BC - CB) \leq \text{rank}(\tilde{B}\tilde{C} - \tilde{C}\tilde{B}) + 2 \leq 2(r-n),$$

and (4.4) is proved.

To establish the theorem let $r \geq n$. We have

$$\begin{aligned} \{t: R(t) \leq r\} &\subset \{t: A \text{ invertible}, n + \tfrac{1}{2} \text{rank}(BA^{-1}C - CA^{-1}B) \leq r\} \\ &\cup \{t: A \text{ not invertible}\}, \end{aligned}$$

and therefore

$$\begin{aligned} \{t: R(t) \leq r\} &\subset \{t: n + \tfrac{1}{2} \text{rank}[B \text{adj}(A)C - C \text{adj}(A)B] \leq r\} \\ &\cup \{t: \text{rank } A < n\}, \end{aligned}$$

where $\text{adj}(A)$ denotes the adjoint of A . Since the right side of this inclusion is closed, the left side may be replaced by its closure $\{t: \underline{R}(t) \leq r\}$. Since this is irreducible, it is contained in one member of the union on the right side. Since $r \geq n$, $\{\underline{R}(t) \leq r\}$ is not contained in $\{\text{rank } A < n\}$. Therefore we have

$$\{t: \underline{R}(t) \leq r\} \subset \{t: n + \tfrac{1}{2} \text{rank}(B \text{adj}(A)C - C \text{adj}(A)B) \leq r\}.$$

So if A is invertible and $r \in \mathbb{N}$ is arbitrary, then

$$\underline{R}(t) \leq r$$

implies $r \geq n$ and hence

$$n + \frac{1}{2} \text{rank}(BA^{-1}C - CA^{-1}B) \leq r.$$

This proves the theorem. ■

Let Λ be a finite dimensional associative algebra (with unity) and N an n -dimensional (unitary, left) Λ -module. The rank $R(N)$ and the border rank $\underline{R}(N)$ of N are defined as $R(t)$ and $\underline{R}(t)$ respectively, where $t \in \Lambda^* \otimes N^* \otimes N$ is the tensor corresponding to the structure map

$$\Lambda \times N \rightarrow N, \quad (a, x) \mapsto ax$$

of the module. In particular, if we take $N = \Lambda$ as left Λ -module, we get the rank $R(\Lambda)$ and the border rank $\underline{R}(\Lambda)$ of the algebra Λ . For $a \in \Lambda$ let $\text{rank}_N(a)$ denote the rank of the linear map

$$\begin{aligned} l_a: N &\rightarrow N, \\ x &\mapsto ax. \end{aligned}$$

COROLLARY 4.2. *Let N be a Λ -module, $\dim N = n$, $b, c \in \Lambda$. Then*

$$\underline{R}(N) \geq n + \frac{1}{2} \text{rank}_N(bc - cb).$$

Proof. W.l.o.g. $1, b, c$ are linearly independent over k . Let

$$U = N^*, \quad V = N, \quad W = (k \cdot 1 + kb + kc)^*.$$

If

$$\pi: \Lambda^* \rightarrow W$$

denotes the dual of the inclusion, the linear map

$$\varphi: \Lambda^* \otimes N^* \otimes N \rightarrow U \otimes V \otimes W$$

$$a^* \otimes x^* \otimes y \mapsto x^* \otimes y \otimes \pi(a^*)$$

sends the structural tensor t of N into a tensor t' , whose slices A, B, C with respect to some basis for V , its dual for U , and the dual of $(1, b, c)$ for W are

the matrices corresponding to l_1, l_b, l_c respectively. In particular $A = \mathbf{1}_n$. So we have

$$\begin{aligned}\underline{R}(N) &= \underline{R}(t) \geq \underline{R}(t') \geq n + \frac{1}{2} \text{rank}(BC - CB) \\ &= n + \frac{1}{2} \text{rank}(l_b l_c - l_c l_b) \\ &= n + \frac{1}{2} \text{rank}_N(bc - cb).\end{aligned}$$

■

As an application, let M_2 be the algebra of 2×2 matrices over k . Then

$$\underline{R}(M_2) \geq 6: \quad (4.6)$$

take

$$b = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad c = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

I do not know whether \geq can be replaced by $=$. Similarly, we get a (possibly sharp) lower bound for the multiplication of a 2×2 matrix with a $2 \times n$ matrix: Let $M_{2 \times n}$ be the M_2 -module of all $2 \times n$ -matrices over k . Then

$$\underline{R}(M_{2 \times n}) \geq 3n. \quad (4.7)$$

(Take b, c as above.) Let T_2 be the algebra of upper triangular 2×2 matrices. Then

$$\underline{R}(T_2) = R(T_2) = 4: \quad (4.8)$$

Take

$$b = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad c = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

More generally, we have $\underline{R}(M_p) \geq \frac{3}{2}p^2$ and $\underline{R}(T_p) \geq p(3p+1)/4$ for $p \geq 2$.

COROLLARY 4.3. *Let $n \geq 3$ be odd. Then the shape $(n, n, 3)$ is balanced, but not good. More precisely, let $r = \lceil 3n^2/(2n+1) \rceil = (3n-1)/2$. Then*

$$\dim X_r \leq 3n^2 - 1.$$

Proof. If $t \in U \otimes V \otimes W$ has slices

$$A = \mathbf{1}_n, \quad B = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}, \quad C = \begin{pmatrix} 0 & 1 & & & \\ & 0 & \cdot & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ 1 & & & & 0 \end{pmatrix}$$

with λ_i pairwise distinct, then

$$\underline{R}(t) \geq n + \frac{1}{2} \text{rank}(BC - CB) = 3n/2;$$

therefore $X_r \neq U \otimes V \otimes W$. ■

The following lemma has independently been proved by T. Lickteig.

LEMMA 4.4. *Let $n \geq 3$ be odd, $\dim U = \dim V = n$, $\dim W = 3$, $r = (3n - 1)/2$. Then*

$$\dim X_r \geq 3n^2 - 1.$$

Proof. It suffices to give $u_1 \otimes v_1 \otimes w_1, \dots, u_r \otimes v_r \otimes w_r \in U \otimes V \otimes W$ such that

$$\dim \sum_{\rho} (U \otimes v_{\rho} \otimes w_{\rho} + u_{\rho} \otimes V \otimes w_{\rho} + u_{\rho} \otimes v_{\rho} \otimes W) \geq 3n^2 - 1.$$

[See (3.4) and (3.5).] Let (e_1, \dots, e_n) , (f_1, \dots, f_n) , and (g_1, g_2, g_3) be bases of U , V , and W respectively, and let $w_1, \dots, w_n, \tilde{w}_3, \tilde{w}_5, \dots, \tilde{w}_n, w \in W$ have coordinates that are algebraically independent over the primefield of k . (By general principles we may assume that k has infinite degree of transcendency.)

Here is our list of the $u_{\rho} \otimes v_{\rho} \otimes w_{\rho}$:

$$\begin{aligned} & e_{\rho} \otimes f_{\rho} \otimes w_{\rho} & (1 \leq \rho \leq n), \\ & (e_{\nu-2} + e_{\nu-1} + e_{\nu}) \otimes (f_{\nu} + f_{\nu+1} + f_{\nu+2}) \otimes \tilde{w}_{\nu} & (3 \leq \nu \leq n-2, \quad \nu \text{ odd}), \\ & (e_{n-2} + e_{n-1} + e_n) \otimes (f_1 + f_2 + f_3) \otimes \tilde{w}_n. \end{aligned}$$

We have to show that $\sum_{\rho} U \otimes v_{\rho} \otimes w_{\rho} + u_{\rho} \otimes V \otimes w_{\rho} + u_{\rho} \otimes v_{\rho} \otimes W$, together with one additional vector (we take $e_1 \otimes f_2 \otimes w$), span all of $U \otimes V \otimes W$. Call a point

$(i, j) \in [1, n] \times [1, n]$ settled iff $e_i \otimes f_j \otimes W$ lies in this span. Now show successively that the following sets consist of settled points (for the ease of presentation nondisjoint sets are used):

$$\begin{aligned} & \{(i, i) : 1 \leq i \leq n\}, \quad \{1, 2, 3\} \times \{1, 2, 3\}, \\ & \{\nu - 2, \nu - 1, \nu\} \times \{\nu, \nu + 1, \nu + 2\}, \quad \{\nu, \nu + 1, \nu + 2\} \times \{\nu, \nu + 1, \nu + 2\} \\ & \hspace{15em} (3 \leq \nu \leq n - 2, \quad \nu \text{ odd}), \\ & \{\nu, \nu + 1, \nu + 2\} \times \{\chi + \nu, \chi + \nu + 1, \chi + \nu + 2\} \\ & \hspace{10em} (4 \leq \chi \leq n - 3, \quad \chi \text{ even}, \quad 1 \leq \nu \leq n - 2 - \chi, \quad \nu \text{ odd}) \end{aligned}$$

[here and below the sets should be taken in lexicographical order with respect to (χ, ν)],

$$\begin{aligned} & \{\chi + \nu, \chi + \nu + 1, \chi + \nu + 2\} \times \{\nu, \nu + 1, \nu + 2\} \\ & (2 \leq \chi \leq n - 3, \quad \chi \text{ even}, \quad 1 \leq \nu \leq n - 2 - \chi, \quad \nu \text{ odd}). \quad \blacksquare \end{aligned}$$

Corollary 4.3 and Lemma 4.4 say that X_r with $r = (3n - 1)/2$ is a hypersurface. We shall now identify the generator of its ideal. Let a_{ij}, b_{ij}, c_{ij} be indeterminates over k ($i, j = 1, \dots, n$). By abuse of notation, put

$$A = (a_{ij}), \quad B = (b_{ij}), \quad C = (c_{ij}).$$

Thus A, B, C are now matrices with indeterminate entries. Consider the rational function

$$F = |A|^2 |BA^{-1}C - CA^{-1}B| \in k(A, B, C). \quad (4.9)$$

($|\cdot|$ denotes the determinant.)

LEMMA 4.5. *Let $n \geq 3$. Then F is an irreducible polynomial.*

Proof. We have

$$F = |A|^2 |C|^2 |C^{-1}BA^{-1} - A^{-1}BC^{-1}| = |C|^2 |AC^{-1}B - BC^{-1}A|.$$

This shows that F is invariant under cyclic permutations of the "slices" A, B, C . Since F is a polynomial in B and C , F is a polynomial. Since F is homogeneous of degree n in C , F is homogeneous of degree n in A and in B as well.

Obviously $F \neq 0$. (See the proof of Corollary 4.3.) Suppose we had a nontrivial decomposition $F = G \cdot H$. Then G and H are homogeneous in each of A, B, C . Substitute for A the identity matrix $\mathbf{1}_n$ and for C a matrix

$$\Lambda = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}, \quad \lambda_i \in k \text{ pairwise distinct.}$$

We get

$$\begin{aligned} G(\mathbf{1}_n, B, \Lambda)H(\mathbf{1}_n, B, \Lambda) &= F(\mathbf{1}_n, B, \Lambda) = |\Lambda B - B\Lambda| \\ &= |((\lambda_i - \lambda_j)b_{ij})| \neq 0. \end{aligned} \quad (4.10)$$

We claim that the latter determinant (call it D) is irreducible. Let $D = D_1 D_2$ be a decomposition. Since D is linear in each row (and each column), the variables of each row (or column) appear in only one of D_1, D_2 . Thus there are partitions

$$[1, n] = P_1 \cup P_2 = Q_1 \cup Q_2, \quad P_1 \cap P_2 = Q_1 \cap Q_2 = \emptyset,$$

such that D_α contains only variables b_{ij} with $(i, j) \in P_\alpha \times Q_\alpha$. But then

$$(P_1 \times Q_1) \cup (P_2 \times Q_2) = \{(i, j) : i \neq j\}.$$

So for $n \geq 3$ the partitions must degenerate, and hence also the decomposition $D = D_1 D_2$.

Since D is irreducible, either $G(\mathbf{1}_n, B, \Lambda)$ or $H(\mathbf{1}_n, B, \Lambda)$ has degree 0 in B . But then also G or H has degree 0 in B . By symmetry the same holds with respect to A and C . So w.l.o.g. G is a polynomial in B , H a polynomial in A and C alone. Let

$$Y = \begin{pmatrix} y_1 & & 0 \\ & \ddots & \\ 0 & & y_n \end{pmatrix}$$

with indeterminates y_i . Then

$$|((y_i - y_j)b_{ij})| = F(\mathbf{1}_n, B, Y) = G(\mathbf{1}_n, B, Y)H(\mathbf{1}_n, B, Y) = \tilde{G}(B)\tilde{H}(Y).$$

Comparing coefficients, we get a contradiction. ■

THEOREM 4.6. *Let $n \geq 3$ be odd, $\dim U = \dim V = n$, $\dim W = 3$, $r = (3n - 1)/2$. Then F generates the ideal of the hypersurface X_r .*

Proof. By Theorem 4.1 and the definition of F we have

$$X_r \cap \{|A| \neq 0\} \subset \{F = 0\}.$$

The left side is nonempty open and therefore dense open in X_r . Hence

$$X_r \subset \{F = 0\}.$$

Here both sides are irreducible hypersurfaces of $U \otimes V \otimes W$. Thus

$$X_r = \{F = 0\}.$$

F is irreducible, so it generates the ideal of X_r . ■

Theorem 4.1 implies $X_r \subset K_r$, where K_r is the closure of the set $\{t: |A| \neq 0, n + \frac{1}{2} \text{rank}(BA^{-1}C - CA^{-1}B) \leq r\}$. Theorem 4.6 implies that for $n \geq 3$, n odd, and $r = (3n - 1)/2$ we have equality. This is an exception. For $k = \mathbb{C}$, $n \geq 7$ odd or even, and $n + 1 \leq r < (3n - 1)/2$ we never have $X_r = K_r$. On the one hand

$$\dim X_r \leq r(2n + 1) =: \xi$$

by (2.5). On the other hand a remark by Hulek [20] implies that K_r is irreducible and

$$\dim K_r = 2n^2 + 4n(r - n) - 4(r - n)^2 =: \eta.$$

Now $\eta - \xi$ is a concave function of r with values $n - 5$ and $(n/2) - 3$ for $r = n + 1$ and $r = (3n/2) - 1$ respectively. Hence $\dim X_r < \dim K_r$ for $n + 1 \leq r < (3n - 1)/2$ and $n \geq 7$.

PROPOSITION 4.7. *Let n be even. Then*

$$\underline{R}(n, n, 3) = 3n/2.$$

Moreover, if $t \in U \otimes V \otimes W$ and $F(t) \neq 0$ then $\underline{R}(t) = 3n/2$.

Proof. $\underline{R}(n, n, 3) \geq 3n/2$ by (2.8). Let $r = 3n/2$. To prove that r is large, use Proposition 3.2 with $u_\rho \otimes v_\rho \otimes w_\rho$ given by

$$\begin{aligned} e_i \otimes f_i \otimes w_i & \quad (1 \leq i \leq n), \\ (e_\nu + e_{\nu+1}) \otimes (f_\nu + f_{\nu+1}) \otimes \tilde{w}_\nu & \quad (1 \leq \nu \leq n-1, \nu \text{ odd}), \end{aligned}$$

where the $w_i, \tilde{w}_\nu \in W$ have algebraically independent coefficients over the primefield of k . (Compare the proof of Lemma 4.4; the present case is much easier, however.) The proof of the second assertion of the proposition is similar to the proof of Theorem 4.6. ■

As an exercise in applying Theorem 4.6 and Proposition 4.7, we shall determine the border rank of an arbitrary finite dimensional \mathfrak{sl}_2 module N for $k = \mathbb{C}$. (As in the associative case, the border rank of N is defined as the border rank of the structural tensor $t \in \mathfrak{sl}_2^* \otimes N^* \otimes N$ of N .) Write

$$N = N(0) \oplus N',$$

where \mathfrak{sl}_2 acts trivially on $N(0)$, and N' is a direct sum of simple \mathfrak{sl}_2 modules $V(m)$ with $m \neq 0$. (See [21, II, 7].) Then obviously

$$\underline{R}(N) = \underline{R}(N').$$

We claim that

$$\underline{R}(N') = \left\lceil \frac{3}{2} \dim N' \right\rceil. \quad (4.11)$$

Let

$$h = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad x = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad y = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$

be the standard basis of \mathfrak{sl}_2 , and A, B, C the matrices with respect to some basis of N' of the action of h, x, y respectively on the module N' . By an

argument similar to the one used in the proof of Corollary 4.2, and by Theorem 4.6 and Proposition 4.7, it suffices to show

$$F(A, B, C) \neq 0.$$

If N' is a direct sum of submodules, F is the product of the corresponding smaller F 's. So we may assume N' simple, i.e., $N' = V(m)$ with $m \neq 0$. Then

$$A = ((m - 2i)\delta_{ij})_{0 \leq i, j \leq m},$$

$$B = ((m - i)\delta_{i, j-1})_{0 \leq i, j \leq m},$$

$$C = (i\delta_{i, j+1})_{0 \leq i, j \leq m}.$$

If m is odd, then A is invertible, and

$$BA^{-1}C - CA^{-1}B = \left(\frac{(m-i)(i+1)}{m-2(i+1)} - \frac{i[m-(i-1)]}{m-2(i-1)} \right)_{i,j} \quad (4.12)$$

is also invertible. So $F(A, B, C) \neq 0$. If m is even, then A is not invertible. To compute $F(A, B, C)$ we use the following device. Replace k by $k(\varepsilon)$, where ε is an indeterminate over k , and replace the zero diagonal term of A by ε . Call this new matrix A_ε . Then A_ε is invertible and $BA_\varepsilon^{-1}C - CA_\varepsilon^{-1}B$ is obtained from the right side of (4.12) by replacing the zero denominators by ε . Thus $F(A_\varepsilon, B, C)$ is a polynomial in ε with nonzero constant term d . Since F is a polynomial we have therefore

$$F(A, B, C) = (F(A_\varepsilon, B, C))_{\varepsilon=0} = d \neq 0.$$

This proves (4.11). In particular we have

$$\underline{R}(\mathfrak{sl}_2) = 5. \quad (4.13)$$

[Lafon (private communication) has shown $R(\mathfrak{sl}_2) = 5$.]

REFERENCES

- 1 A. Alder, Dissertation, Universität Zürich, 1983.
- 2 A. Alder and V. Strassen, On the algorithmic complexity of associative algebras, *Theoret. Comput. Sci.* 15:201-211 (1981).
- 3 M. D. Atkinson and S. Lloyd, Bounds on the ranks of some 3-tensors, *Linear Algebra Appl.* 31:19-31 (1980).

- 4 M. D. Atkinson and N. M. Stephens, On the maximal multiplicative complexity of a family of bilinear forms, *Linear Algebra Appl.* 27:1-8 (1979).
- 5 D. Bini, Relations between EC-algorithms and APA-algorithms, applications, *Calcolo* 17:87-97 (1980).
- 6 D. Bini, M. Capovani, G. Lotti, and F. Romani, $O(n^{2.7799})$ complexity for matrix multiplication, *Inform. Process. Lett.* 8:234-235 (1979).
- 7 A. Borodin and I. Munro, *Computational Complexity of Algebraic and Numeric Problems*, American Elsevier, New York, 1975.
- 8 R. W. Brockett, On the generic degree of a 3-tensor, unpublished manuscript, Harvard Univ., 1976.
- 9 R. W. Brockett and D. Dobkin, On the optimal evaluation of a set of bilinear forms, in *Proceedings of the 5th Annual ACM Symposium on the Theory of Computing*, 1973, pp. 88-95.
- 10 D. Dobkin, On the optimal evaluation of a set of n -linear forms, in *Conference Record, 14th Annual Symposium on Switching and Automata Theory*, Iowa City, 1973, pp. 92-102.
- 11 C. M. Fiduccia, On obtaining upper bounds on the complexity of matrix multiplication, in *Complexity of Computer Computations* (R. Miller and J. Thatcher, Eds.), Plenum, New York, 1972.
- 12 N. Gastinel, Sur le calcul des produits de matrices, *Numer. Math.* 17:222-229 (1971).
- 13 D. Yu. Grigoryev, Some new bounds on tensor rank, LOMI preprint E-2-1978, Leningrad, 1978.
- 14 H. F. de Groote, On varieties of optimal algorithms for the computation of bilinear mappings. I. The isotropy group of a bilinear mapping, *Theoret. Comput. Sci.* 7:1-24 (1978).
- 15 H. F. de Groote, On varieties of optimal algorithms for the computation of bilinear mappings. II. Optimal algorithms for 2×2 -matrix multiplication, *Theoret. Comput. Sci.* 7:127-148 (1978).
- 16 H. F. de Groote, On varieties of optimal algorithms for the computation of bilinear mappings. III. Optimal algorithms for the computation of xy and yx where $x, y \in M_2(K)$.
- 17 H. F. de Groote, Characterization of division algebras of minimal ranks and the structure of their algorithm varieties, *SIAM J. Comput.*, to appear.
- 18 R. Hartshorne, *Algebraic Geometry*, Graduate Texts in Math., No. 52, Springer, New York, 1977.
- 19 T. D. Howell, Global properties of tensor rank, *Linear Algebra Appl.* 22:9-23 (1978).
- 20 K. Hulek, A remark on certain matrix varieties, Preprint, Universität Erlangen, 1981.
- 21 J. E. Humphreys, *Introduction to Lie Algebras and Representation Theory*, Springer, New York, 1972.
- 22 J. Ja'Ja, Optimal evaluation of pairs of bilinear forms, in *Proceedings, 10th Annual ACM Symposium on Theory of Computing*, 1978, pp. 173-183.

- 23 J. B. Kruskal, Trilinear decomposition of three-way arrays: rank and uniqueness in arithmetic complexity and statistical models, *Linear Algebra Appl.* 18:95–138 (1977).
- 24 J. C. Lafon, Base tensorielle des matrices de Hankel (ou de Toeplitz) applications, *Numer. Math.*, 1975, pp. 349–361.
- 25 J. C. Lafon, Complexité de l'évaluation de plusieurs formes bilinéaires et des principaux calculs matriciels, Thèse, Grenoble, Nov. 1976.
- 26 D. Mumford, *Algebraic Geometry I*, Grundlehren der math. Wissensch., No. 221, Springer, Berlin, 1976.
- 27 A. M. Ostrowski, On two problems in abstract algebra connected with Horner's rule, in *Studies in Mathematics and Mechanics presented to Richard von Mises*, Academic, New York, 1954, pp. 40–48.
- 28 I. R. Shafarevich, *Grundzüge der Algebraischen Geometrie*, Vieweg, Braunschweig, 1972.
- 29 I. R. Shafarevich, *Basic Algebraic Geometry*, Grundlehren der math. Wissensch., No. 213, Springer, Berlin, 1974.
- 30 V. Strassen, Evaluation of rational functions, in *Complexity of Computer Computations* (R. Miller and J. Thatcher, Eds.), Plenum, New York, 1972, pp. 1–10.
- 31 V. Strassen, Vermeidung von Divisionen, *Crelles J. Reine Angew. Math.* 264:184–202 (1973).
- 32 S. Winograd, Some bilinear forms whose multiplicative complexity depends on the field of constants, *Math. Systems Theory* 10:169–180 (1977).

Received 3 May 1982