

TENSOR RANK, INVARIANTS, INEQUALITIES, AND APPLICATIONS*

ELIZABETH S. ALLMAN[†], PETER D. JARVIS[‡], JOHN A. RHODES[†], AND
JEREMY G. SUMNER[‡]

Abstract. Though algebraic geometry over \mathbb{C} is often used to describe the closure of the tensors of a given size and complex rank, this variety includes tensors of both smaller and larger rank. Here we focus on the $n \times n \times n$ tensors of rank n over \mathbb{C} , which has as a dense subset the orbit of a single tensor under a natural group action. We construct polynomial invariants under this group action whose nonvanishing distinguishes this orbit from points only in its closure. Together with an explicit subset of the defining polynomials of the variety, this gives a semialgebraic description of the tensors of rank n and multilinear rank (n, n, n) . The polynomials we construct coincide with Cayley’s hyperdeterminant in the case $n = 2$ and thus generalize it. Though our construction is direct and explicit, we also recast our functions in the language of representation theory for additional insights. We give three applications in different directions: First, we develop basic topological understanding of how the real tensors of complex rank n and multilinear rank (n, n, n) form a collection of path-connected subsets, one of which contains tensors of real rank n . Second, we use the invariants to develop a semialgebraic description of the set of probability distributions that can arise from a simple stochastic model with a hidden variable, a model that is important in phylogenetics and other fields. Third, we construct simple examples of tensors of rank $2n - 1$ which lie in the closure of those of rank n .

Key words. tensor rank, hyperdeterminant, border rank, latent class model, phylogenetics

AMS subject classifications. 15A72, 14P10

DOI. 10.1137/120899066

1. Introduction. The notion of tensor rank naturally extends the familiar notion of matrix rank for two-dimensional numerical arrays to d -dimensional arrays and likewise has extensive connections to applied problems. However, basic questions about tensor rank can be much more difficult to answer than their matrix analogues, and many open problems remain. Several natural problems concerning tensor rank are to determine for a given field the rank of an explicitly given tensor, to determine for a given field and format the possible ranks of all tensors, and to determine for a given field and format the generic rank(s) of a tensor. While the matrix versions of these problems are solved by an understanding of Gaussian elimination and determinants, for higher-dimensional tensors they have so far eluded general solutions.

The case of $2 \times 2 \times 2$ tensors, however, is quite well studied [10] and provides one model of desirable understanding: Over \mathbb{C} or \mathbb{R} , such a tensor may have rank 0, 1, 2, or 3 only. Over \mathbb{C} , descriptions of the sets of tensors of each of these possible ranks may be given, in terms of intersections, unions, and complements of explicit algebraic varieties. These descriptions can thus be phrased as Boolean combinations

*Received by the editors November 15, 2012; accepted for publication (in revised form) by L. De Lathauwer April 24, 2013; published electronically July 11, 2013.

<http://www.siam.org/journals/simax/34-3/89906.html>

[†]Department of Mathematics and Statistics, University of Alaska, Fairbanks, AK 99775 (e.allman@alaska.edu, j.rhodes@alaska.edu). The work of these authors was supported in part by U.S. National Science Foundation grant DMS 0714830, the University of Tasmania Visiting Scholar Program, and the Institut Mittag-Leffler.

[‡]School of Mathematics and Physics, University of Tasmania, Hobart, TAS 7001, Australia (peter.jarvis@utas.edu.au, jsumner@utas.edu.au). The work of these authors was supported in part by Australian Research Council grants DP0877447 and FT100100031.

of polynomial equalities. Over \mathbb{R} , analogous explicit descriptions require polynomial inequalities using “>” as well, and the descriptions are thus semialgebraic. For larger tensors of any given rank the existence of a semialgebraic description is a consequence of the Tarski–Seidenberg theorem [34, 30]. While in theory quantifier elimination can produce such a description, in practice the procedure is infeasible, and complete explicit descriptions are generally not known.

A polynomial of particular importance in understanding the $2 \times 2 \times 2$ case is Cayley’s hyperdeterminant [8],

$$\begin{aligned}
 (1.1) \quad \Delta(P) = & (p_{000}^2 p_{111}^2 + p_{001}^2 p_{110}^2 + p_{010}^2 p_{101}^2 + p_{011}^2 p_{100}^2) \\
 & - 2(p_{000} p_{001} p_{110} p_{111} + p_{000} p_{010} p_{101} p_{111} + p_{000} p_{011} p_{100} p_{111} \\
 & \quad + p_{001} p_{010} p_{101} p_{110} + p_{001} p_{011} p_{110} p_{100} + p_{010} p_{011} p_{101} p_{100}) \\
 & + 4(p_{000} p_{011} p_{101} p_{110} + p_{001} p_{010} p_{100} p_{111}),
 \end{aligned}$$

also known as the *tangle* in the physics literature [9]. The function Δ has nonzero values precisely on a certain dense subset of those $2 \times 2 \times 2$ tensors of complex tensor rank 2. If a tensor is real, the sign of the Δ further indicates information about its tensor rank over \mathbb{R} [10]: If $\Delta > 0$, then the tensor has real tensor rank 2, and if $\Delta < 0$, its real tensor rank is 3.

The role of Δ here can be partially understood as a consequence of it being an invariant of the group $GL(2, \mathbb{C}) \times GL(2, \mathbb{C}) \times GL(2, \mathbb{C})$, which acts on $2 \times 2 \times 2$ complex tensors in the three indices and preserves their tensor rank. The transformation property of Δ under this group, along with explicit evaluation at a particular tensor of rank 2, implies that it is nonzero on an orbit which is dense among all tensors of rank 2. The fact that it is zero off of this orbit can be shown by first determining a list of canonical representatives of other orbits and then explicitly evaluating Δ on them to see that it vanishes. Thus both the transformation property of Δ under the group and the ability to evaluate Δ at specific points are essential.

In this work we focus on $n \times n \times n$ tensors of tensor rank n , over \mathbb{C} and over \mathbb{R} , with the goal of generalizing our detailed understanding of the $2 \times 2 \times 2$ tensors to this particular case. Although we do not translate our results here to the cases of $n_1 \times n_2 \times n_3$ tensors of rank n with $n \leq n_i$, this should be possible by applying maps of $\mathbb{C}^{n_i} \rightarrow \mathbb{C}^n$. Thus what is most important about this case is that the dimensions of the tensor are sufficiently large that they do not put constrictions on studying the given rank n . (A more careful reading will show that for many arguments we only need $n_i \geq n$ for at least two values of i .)

Over \mathbb{C} , the rank-2 tensors are dense among all $2 \times 2 \times 2$ tensors. However, for $n > 2$, the closure of the rank- n tensors in the $n \times n \times n$ ones forms an algebraic variety of dimension strictly less than n^3 . From the closure operation, this variety contains all tensors of rank $< n$, as well as some of rank $> n$. Much previous work has focused on determining defining polynomials of this variety, that is, polynomials that vanish on all such rank- n tensors. For $n = 3$, the ideal of polynomials defining this variety is known [15]. For $n = 4$ a set-theoretic defining set of polynomials has been determined [12, 7, 13]. For all $n \geq 3$, many polynomials in the ideal are known through a general construction of “commutation relations” [3, 4] (see also [31]). Moreover, the commutation relations give the full ideal up to an explicit saturation, and taking a radical. Nonetheless, the full ideal is still not understood if $n \geq 4$.

In this paper we turn from studying polynomial equalities related to tensor rank issues to inequalities. Our main contribution is a generalization to arbitrary n of the $n = 2$ hyperdeterminant, Δ , of Cayley. We obtain polynomial functions whose

nonvanishing singles out a dense orbit of the tensors of rank n from their closure. We emphasize that this generalization does *not* lead to those functions standardly called “hyperdeterminants” in the modern mathematics literature [16], but rather to a set of functions that generalize the properties of Δ in another way, appropriate to the problem at hand. Just as Δ defines a one-dimensional representation of $GL(2, \mathbb{C}) \times GL(2, \mathbb{C}) \times GL(2, \mathbb{C})$, our functions determine a multidimensional representation of $GL(n, \mathbb{C}) \times GL(n, \mathbb{C}) \times GL(n, \mathbb{C})$.

The usefulness of studying invariant spaces of polynomials for investigating tensor rank issues is, of course, not new (see, for instance, the survey [23] and its references for instances going beyond Δ). Because the relevant representation theory is so highly developed, such study can be fairly abstract. While the gap between abstractly understanding such polynomials and concretely evaluating them is conceptually a small one, in practice it is by no means trivially bridged ([7] gives an excellent illustration of this). Since our arguments depend crucially on being able to explicitly evaluate our invariant polynomials on certain tensors, a concrete approach to developing them is warranted here.

Specifically, we investigate the transformation properties of our functions under a group action, a reduction under that action of most tensors to a semicanonical form which is made possible by knowledge of the commutation relations, and the evaluation of our functions at these semicanonical forms. Together, these allow us to make precise statements about the zero set of these polynomials within the closure of the rank- n $n \times n \times n$ tensors that are analogous to statements about the zero set of Δ in the $2 \times 2 \times 2$ case. We also relate the functions we construct to certain others with similar transformation properties and show that those can be used as substitutes, provided that they do not vanish at a point. This leads us to reframe our work in the language of representation theory. Finally we show how our generalization of Δ can be used for three different applications.

This paper is organized as follows: After definitions and preliminaries in section 2, in section 3 we recall relevant facts about the algebraic variety of $n \times n \times n$ tensors of rank n and construct semicanonical orbit representatives under the group action. We then construct our invariant polynomial functions in section 4 and determine on which complex tensors of complex border rank at most n they vanish. Then in section 5 we study these functions in the framework of representation theory for the group $GL(n, \mathbb{C}) \times GL(n, \mathbb{C}) \times GL(n, \mathbb{C})$.

As a first application, in section 6 we use these polynomials to investigate real tensors of complex rank n . We show that the zero set of our invariants divides the real points on the variety into several path components, each of which contains tensors of a single *signature*. These signatures are characterized by the number of complex conjugate pairs of rank-1 tensors in their unique rank- n decomposition. We also determine the number of path components of each signature and find that the tensors of real rank n form a single component. We extend, from $n = 2$ to $n = 3$, the result that the sign of a polynomial invariant distinguishes whether a tensor of complex rank n also has real rank n . For larger n , the sign of our invariant is insufficient to single out the component composed of tensors of real rank n , but why it fails to do so is made clear.

In section 7 we turn to the application which originally motivated our interest in $n \times n \times n$ tensors of rank n , which is their appearance as certain statistical models, in both latent class analysis and phylogenetics. For these applications (which we introduce more thoroughly in section 7), such a tensor represents a joint probability distribution of three observed random variables, each with discrete state space of

size n , and thus has nonnegative entries summing to 1. Its decomposition into a sum of rank-1 tensors reflects the structure of the stochastic model, in which the distributions of each of the observed variables depend on the state of a common hidden (latent, or unobservable) variable with n states. The phylogenetic application can be seen through an interpretation of the observed variables as having four states, the bases A, C, G, T that may appear at a particular site in DNA sequences from three species, while the state of the hidden variable represents the base in an ancestral organism from which the others evolved.

For these statistical applications, the rank-1 components of these tensors are themselves probability distributions, up to scaling, so it is important to not only determine the rank of a tensor, but also be able to determine whether the rank-1 components have nonnegative entries. The $2 \times 2 \times 2$ rank-2 case and its extension to phylogenetic trees have recently been studied in [36] (see also [19, 28]). In that work Cayley’s hyperdeterminant Δ played an important role. Our work here began as a step toward extending some of the results of [36] from $n = 2$ to $n > 2$. In this paper, however, we limit ourselves to the simplest phylogenetic model (on a 3-leaf tree), as the extension to larger trees depends on other ideas which will be presented in [5]. Despite lacking a good test for determining that a tensor of complex rank n has real rank n , we borrow ideas from [5] to give semialgebraic conditions that ensure a tensor is a probability distribution arising from the latent class model (with certain mild conditions on the parameters).

As a final application of the main theorems of this paper, in section 8 we show examples of $n \times n \times n$ tensors of border rank n , but rank larger than n . We give a simple, explicit example of a tensor of this sort with rank $2n - 1$. When $n = 2$ this gives the well-known canonical form of a complex rank-3 tensor; however, for general $n > 2$ this produces a new class of examples of “rank jumping” by a large amount.

2. Three-dimensional tensors, group actions, and rank. Denote the space of all complex tensors of *format* (n_1, n_2, n_3) by $S = S(n_1, n_2, n_3) = \mathbb{C}^{n_1} \otimes \mathbb{C}^{n_2} \otimes \mathbb{C}^{n_3}$. Note that $S \cong \mathbb{C}^{n_1 n_2 n_3}$, but that one may view an element of S concretely as an $n_1 \times n_2 \times n_3$ array of complex numbers. We thus identify such tensors with three-dimensional hypermatrices.

Let

$$G(\mathbb{C}) = GL(n_1, \mathbb{C}) \times GL(n_2, \mathbb{C}) \times GL(n_3, \mathbb{C})$$

and

$$G(\mathbb{R}) = GL(n_1, \mathbb{R}) \times GL(n_2, \mathbb{R}) \times GL(n_3, \mathbb{R}) \subset G(\mathbb{C}),$$

which act on S through the three indices of tensors. We write this action on the right, using several interchangeable notations, so that for $P \in S, (g_1, g_2, g_3) \in G(\mathbb{C})$,

$$P(g_1, g_2, g_3) = ((P *_1 g_1) *_2 g_2) *_3 g_3 = \cdots = ((P *_3 g_3) *_2 g_2) *_1 g_1,$$

where, for instance,

$$(P *_3 g_3)_{ijk} = \sum_{l=1}^{n_3} P_{ijl} g_3(l, k),$$

with similar formulas for the action in other indices. This notation is also useful for multiplication by vectors, so that if $\mathbf{v} \in \mathbb{C}^{n_3}$, for instance, then $P *_3 \mathbf{v}$ is a matrix

with entries

$$(P *_3 \mathbf{v})_{ij} = \sum_{k=1}^{n_3} P_{ijk} v_k.$$

For $i \in \{1, 2, 3\}$, by the i -slices of P we mean the n_i matrices $P *_i \mathbf{e}_j$, whose entries have the i th index fixed as j . For example, for $i = 3$, a tensor P has matrix slices P_1, P_2, \dots, P_{n_3} , where $P_j = P(\cdot, \cdot, j)$. The action of $G(\mathbb{C})$ on a tensor can be understood through transformation of these slices, as $P' = P(g_1, g_2, I)$ has slices $P'_j = g_1^T P_j g_2$, while $P'' = P(I, I, g_3)$ has slices $P''_j = \sum_k P_k g_3(k, j)$.

The *complex tensor rank* of $P \in S$ is the smallest integer r such that

$$P = \sum_{i=1}^r \mathbf{u}_i \otimes \mathbf{v}_i \otimes \mathbf{w}_i$$

for some $\mathbf{u}_i \in \mathbb{C}^{n_1}$, $\mathbf{v}_i \in \mathbb{C}^{n_2}$, and $\mathbf{w}_i \in \mathbb{C}^{n_3}$. For a real tensor, the real tensor rank is defined analogously, requiring that $\mathbf{u}_i, \mathbf{v}_i, \mathbf{w}_i$ be real. Note that if a tensor is real, its real and complex tensor ranks need not be equal, though the complex tensor rank is an obvious lower bound for the real tensor rank.

There is also a notion of *multilinear rank* of such a tensor, which is an ordered triple (r_1, r_2, r_3) . Here r_i is the rank of the transformation

$$\begin{aligned} \mathbb{C}^{n_i} &\rightarrow \mathbb{C}^{n_j} \otimes \mathbb{C}^{n_k}, \\ \mathbf{v} &\mapsto P *_i \mathbf{v} \end{aligned}$$

associated to P , and thus is the ordinary matrix rank of the $n_j n_k \times n_i$ flattening of P . The multilinear rank of a real tensor is thus independent of the choice of field \mathbb{R} or \mathbb{C} , as the analogous fact holds for matrices.

Both tensor rank (which we often will refer to as simply *rank*, or \mathbb{C} -rank or \mathbb{R} -rank if the field must be made clear) and multilinear rank are invariant under the action of the general linear group. More precisely, if $P' = Pg$ with $g \in G(\mathbb{C})$, then P' and P have the same \mathbb{C} -rank and multilinear rank. If P is real and $g \in G(\mathbb{R})$, then P and P' have the same \mathbb{R} -rank as well.

For the remainder of the paper, we restrict our attention to the case

$$n_1 = n_2 = n_3 = n,$$

although, as pointed out in the introduction, many results are easily modified to cases with $n_i \geq n$.

If $\mathbf{v} \in \mathbb{C}^n$, let $\text{diag}(\mathbf{v})$ denote the $n \times n$ diagonal matrix whose (i, i) -entry is v_i . Similarly, let $\text{Diag}(\mathbf{v})$ denote the $n \times n \times n$ diagonal tensor with (i, i, i) -entry v_i . In particular, if $\mathbf{1}$ is the vector with all entries 1, then $\text{diag}(\mathbf{1}) = I_n$ is the identity matrix. We denote its tensor analogue by $D = D_n = \text{Diag}(\mathbf{1})$.

By $\mathcal{D}(\mathbb{C})$ and $\mathcal{D}(\mathbb{R})$ we denote the $G(\mathbb{C})$ - and $G(\mathbb{R})$ -orbits of D , respectively.

PROPOSITION 2.1. $\mathcal{D}(\mathbb{C})$ is the set of all $n \times n \times n$ complex tensors of \mathbb{C} -rank n and multilinear rank (n, n, n) .

$\mathcal{D}(\mathbb{R})$ is the set of all $n \times n \times n$ real tensors of \mathbb{R} -rank n and multilinear rank (n, n, n) .

Proof. First, observe that $D = \text{Diag}(\mathbf{1})$ has \mathbb{C} -rank and \mathbb{R} -rank n and multilinear rank (n, n, n) : The multilinear rank is clear since the $n^2 \times n$ flattenings of D all have

the standard basis vectors among their rows. The \mathbb{C} -rank and \mathbb{R} -rank of D are at most n , since

$$D = \sum_{i=1}^n \mathbf{e}_i \otimes \mathbf{e}_i \otimes \mathbf{e}_i.$$

If D had \mathbb{C} -rank or \mathbb{R} -rank $k < n$, so that $D = \sum_{i=1}^k \mathbf{u}_i \otimes \mathbf{v}_i \otimes \mathbf{w}_i$, then

$$I_n = D *_3 \mathbf{1} = \sum_{i=1}^k (\mathbf{w}_i \cdot \mathbf{1}) \mathbf{u}_i \otimes \mathbf{v}_i$$

would have matrix rank less than n , which is absurd.

That every element of these orbits has the stated tensor rank and multilinear rank is a consequence of their invariance under the group actions.

To see that every complex tensor P of rank n and multilinear rank (n, n, n) lies in $\mathcal{D}(\mathbb{C})$, first write

$$P = \sum_{i=1}^n \mathbf{u}_i \otimes \mathbf{v}_i \otimes \mathbf{w}_i.$$

Then since the first flattening of P to an $n^2 \times n$ matrix has rank n , the \mathbf{u}_i must be independent, so the matrix g_1 with i th row \mathbf{u}_i is in $GL(n, \mathbb{C})$. By the same reasoning, the matrices g_2, g_3 with i th rows $\mathbf{v}_i, \mathbf{w}_i$ are in $GL(n, \mathbb{C})$. Then one checks that $P = D(g_1, g_2, g_3) \in \mathcal{D}(\mathbb{C})$. The same argument applies in the real case. \square

Note that not every tensor of \mathbb{C} -rank n is in $\mathcal{D}(\mathbb{C})$. For instance, the tensor P with slices $I, 0, 0, \dots, 0$ has tensor rank n , but is not in $\mathcal{D}(\mathbb{C})$ since it has multilinear rank $(n, n, 1)$. However, P is in the closure of $\mathcal{D}(\mathbb{C})$, since one can give a sequence $\{h_i\}$ of elements in $GL(n, \mathbb{C})$ with $\lim h_i = \mathbf{1e}_1^T$, and then $\lim D(I, I, h_i) = P$. Similar reasoning shows that any tensor of complex tensor rank $\leq n$ is in the closure of $\mathcal{D}(\mathbb{C})$ and that an analogous statement holds for $\mathcal{D}(\mathbb{R})$.

Finally, note that the closures of $\mathcal{D}(\mathbb{C})$ and $\mathcal{D}(\mathbb{R})$ also contain tensors of rank greater than n . Indeed, the phenomenon that tensor rank may *increase* when one takes a limit is a key difference from the matrix rank. We will return to this with some explicit examples in section 8.

3. The variety of rank- n tensors, and certain orbit representatives. Let $V_n \subseteq \mathbb{C}^{n^3}$ denote the closure of $\mathcal{D}(\mathbb{C})$, under either the Zariski or standard topology, as these give the same set. This is the smallest algebraic variety containing all tensors of \mathbb{C} -rank n and multilinear rank (n, n, n) . (It is straightforward to see that it is also the smallest variety containing all tensors of \mathbb{C} -rank n .) Since V_n is the closure of a $G(\mathbb{C})$ -invariant set, V_n is also $G(\mathbb{C})$ -invariant. As mentioned in section 2, for all $n \geq 2$, V_n contains both tensors of rank less than n and tensors of rank greater than n . Tensors in $V_n \setminus V_{n-1}$ have rank $\geq n$, and are said to have *border rank* n .

A key fact we will use is that for all n some defining equations for the variety V_n are known, those given by the *commutation relations* [3, 4]. The essential idea behind these first appeared in the literature on tensors in [31], though much earlier work in the statistics literature on latent class models contains the idea as well.

PROPOSITION 3.1. *The ideal $I(V_n)$ of polynomials vanishing on V_n includes those obtained from entries of the following matrix equations, $i = 1, 2, 3$, $1 \leq j < k \leq n$:*

$$(3.1) \quad (P *_i \mathbf{e}_j) \operatorname{adj}(P *_i \mathbf{v})(P *_i \mathbf{e}_k) - (P *_i \mathbf{e}_k) \operatorname{adj}(P *_i \mathbf{v})(P *_i \mathbf{e}_j) = 0,$$

where $\mathbf{v} \in \mathbb{C}^n$ is an arbitrary vector, and adj denotes the classical adjoint matrix.

While the identity (3.1) still holds if the $\mathbf{e}_j, \mathbf{e}_k$ appearing in it are replaced by more general vectors, this yields only linear combinations of the identities above and thus no essentially new polynomials. Moreover, by treating \mathbf{v} as a vector of indeterminates and considering the coefficients of the monomials in \mathbf{v} that result from (3.1), one may list a finite set of polynomials in P linearly spanning this set for all choices of \mathbf{v} .

In the case $n = 2$, these relations are all trivial (i.e., simplify to $0 = 0$). If $n = 3$, these polynomials are known to generate $I(V_3)$ [15]. For $n \geq 4$, it is known that additional polynomials are needed to generate $I(V_n)$, but little is known about them. A reward offered by one of the authors (ESA) in 2007 for determining $I(V_4)$ has led to that question being called “the salmon problem.” Currently, only set-theoretic defining polynomials have been determined [12, 7, 13].

Though the orbit of D is dense in V_n , additional orbits lie in V_n as well. Next we show that some of the $G(\mathbb{C})$ -orbits in V_n have orbit representatives of a certain form. This semicanonical form will be used for determining on which tensors the functions constructed in the next section vanish. While the determination of orbit representative for tensors of certain specific sizes has been studied by many (e.g., [26]) and usually leads to long lists of canonical forms, we can avoid addressing the general problem as we are focused solely on tensors of a specific rank and do not need our forms to be unique within an orbit.

DEFINITION 3.2. *An $n \times n \times n$ tensor P is i -slice-nonsingular if there is some \mathbb{C} -linear combination of the i -slices that is nonsingular, and slice-nonsingular if it is i -slice-nonsingular for some $i \in \{1, 2, 3\}$. If P is not i -slice-nonsingular, we say it is i -slice-singular. If P is not slice-nonsingular, we say it is slice-singular.*

Note that P is i -slice-nonsingular if and only if one can act on P in the i th index by an element of $GL(n, \mathbb{C})$ to obtain a tensor with a nonsingular i -slice. Thus the terms in the above definition all depend only on the $G(\mathbb{C})$ -orbit of P . To investigate orbits, we consider slice-singular and slice-nonsingular orbits separately.

Let $\mathbf{x} = (x_1, x_2, \dots, x_n)$ be a vector of indeterminates. Then P is i -slice-singular precisely when $h_i(P; \mathbf{x}) = \det(P *_i \mathbf{x})$ is the zero polynomial in \mathbf{x} . Thus the i -slice-singular tensors form an algebraic variety, defined by setting equal to zero the coefficients of each \mathbf{x} -monomial in the expansion of h_i .

PROPOSITION 3.3. *Suppose P is i -slice-nonsingular and for that i satisfies the polynomials of (3.1) in Proposition 3.1. Then P has a $G(\mathbb{C})$ -orbit representative with all i -slices upper triangular.*

Moreover, if the matrix Z whose columns are the diagonals of the i -slices of such a representative is nonsingular, then $P \in \mathcal{D}(\mathbb{C})$. If Z is singular, then an orbit representative exists for P with upper triangular i -slices and at least one slice strictly upper triangular.

Theorem 4.1 below implies that the matrix Z of this theorem is nonsingular precisely when $P \in \mathcal{D}(\mathbb{C})$.

Proof. For convenience, suppose P is 3-slice-nonsingular, with 3-slices P_1, P_2, \dots, P_n . Then, passing to other elements in its $G(\mathbb{C})$ -orbit, we may first assume that P has a nonsingular slice, and then that it has an identity slice, say $P_1 = I$. But then the commutation relations of Proposition 3.1 with $\mathbf{v} = \mathbf{e}_1$ require that for any

$1 \leq j, k \leq n,$

$$P_j \operatorname{adj}(P_1) P_k - P_k \operatorname{adj}(P_1) P_j = 0,$$

so

$$P_j P_k = P_k P_j.$$

Since these parallel slices commute, they can be simultaneously upper-triangularized, by a unitary g_1 . Thus, acting by $(g_1^T, g_1^{-1}, I) \in G(\mathbb{C})$, we may assume that the slices P_i are all upper-triangular.

Let Z be the matrix whose columns are the diagonals of the slices; i.e., $Z(i, j) = P(i, i, j)$.

Suppose first that Z is nonsingular. Then acting on P by (I, I, g_2) for appropriately chosen g_2 will preserve the upper-triangular form of the slices, keeping $P_1 = I$, but making another slice, say P_2 , have distinct entries on its diagonal: To see this, note that the action of (I, I, g_2) on P will send Z to Zg_2 . Choose some nonsingular Z' whose first column is $\mathbf{1}$ and whose second column has distinct entries, and choose g_2 so $Zg_2 = Z'$. Then $P' = P(I, I, g_2)$ will have all upper-triangular slices, with $\mathbf{1}$ on the diagonal of P'_1 and distinct entries on the diagonal of P'_2 . To see that in fact $P'_1 = I$, for any fixed $i < j$ consider the row vector $\mathbf{w}_{ij} = P(i, j, \cdot)$, whose entries come from the strictly upper-triangular entries of the 3-slices. Now there is some row vector \mathbf{a} with $\mathbf{w}_{ij} = \mathbf{a}Z$. But since the first entry of \mathbf{w}_{ij} is 0, and the first column of Z is $\mathbf{1}$, we see $0 = \mathbf{a}\mathbf{1}$. Thus $\mathbf{w}_{ij}g_2 = \mathbf{a}Zg_2 = \mathbf{a}Z'$ implies that the first entry of $\mathbf{w}_{ij}g_2$ is also 0, since the first column of Z' is $\mathbf{1}$.

Assuming now that P_2 has distinct entries on its diagonal, it can be diagonalized by acting on P by some (g_3^T, g_3^{-1}, I) , without changing $P_1 = I$ or the diagonal entries of P_2 . But then the commutation of P_2 with all other slices shows that they are also diagonal. Moreover, Z being nonsingular is equivalent to a statement that no nonzero linear combination of the upper-triangular slices is nilpotent. This property is preserved by the action of (g_3^T, g_3^{-1}, I) , and so the new matrix Z of diagonals is nonsingular as well. Thus, by a final action by (I, I, Z^{-1}) , we obtain D .

If, on the other hand, Z is singular when the P_i are upper-triangular, then there exists a $g_4 \in GL(n, \mathbb{C})$ with Zg_4 having a column of zeros. Acting on P by (I, I, g_4) preserves the upper-triangular form of the slices but ensures that one slice has zeros on the diagonal. \square

4. Construction of invariant functions, and their behavior on V_n . In this section, for all $n > 2$ we construct explicit polynomial and rational functions on the $n \times n \times n$ tensors, with invariance properties under $G(\mathbb{C})$. When $n = 2$ this construction gives Cayley’s hyperdeterminant Δ , though for larger n we were not able to find this generalization in the literature. We then investigate the values these functions take on when restricted to V_n . Using Proposition 3.3, the explicitness of our construction allows us to show that the nonvanishing of the functions distinguishes the orbit $\mathcal{D}(\mathbb{C})$.

Let $\mathbf{x} = (x_1, \dots, x_n)$ be a column vector of auxiliary indeterminates. For an $n \times n \times n$ tensor P , consider the following functions for $i \in \{1, 2, 3\}$:

$$(4.1) \quad h_i(P; \mathbf{x}) = \det(P *_i \mathbf{x}),$$

$$(4.2) \quad f_i(P; \mathbf{x}) = (-1)^{n-1} \det(H_{\mathbf{x}}(h_i(P; \mathbf{x}))).$$

Here $\det M$ denotes the determinant of a matrix M , and $H_{\mathbf{x}}$ is the Hessian operator on a scalar-valued function, giving the matrix of second-order partial derivatives with respect to the indeterminates \mathbf{x} .

These functions are polynomials in the entries of P and \mathbf{x} , homogeneous in each. Their degrees are

$$(4.3) \quad \deg_P(h_i) = n, \quad \deg_{\mathbf{x}}(h_i) = n,$$

$$(4.4) \quad \deg_P(f_i) = n^2, \quad \deg_{\mathbf{x}}(f_i) = n(n - 2).$$

From the action of $G(\mathbb{C})$ on tensors, the functions above inherit certain invariance properties. If $P' = P(g_1, g_2, g_3)$ and $\{i, j, k\} = \{1, 2, 3\}$, then one sees

$$(4.5) \quad h_i(P'; \mathbf{x}) = \det(g_j) \det(g_k) h_i(P; g_i \mathbf{x}).$$

Since, by the chain rule,

$$H_{\mathbf{x}}(h_i(P'; \mathbf{x})) = \det(g_j) \det(g_k) g_i^T ((H_{\mathbf{x}} h_i)(P; g_i \mathbf{x})) g_i,$$

taking determinants yields

$$(4.6) \quad f_i(P'; \mathbf{x}) = \det(g_j)^n \det(g_k)^n \det(g_i)^2 f_i(P; g_i \mathbf{x}).$$

Remark. When $n = 2$, note that $f_i(P; \mathbf{x}) = f_i(P)$ is independent of \mathbf{x} and can be seen to be independent of i as well, by calculating its explicit formula. Moreover, $f_i(P) = \Delta(P)$ since in this case our construction is exactly Schläfli's construction of the $2 \times 2 \times 2$ hyperdeterminant from the 2×2 determinant: Since $\det(P *_i \mathbf{x})$ is a quadratic form when $n = 2$, the determinant of the Hessian of this form is the same as the discriminant of the form. Schläfli's construction is typically presented using the discriminant [16], and that formulation then generalizes to yield (a multiple of) the hyperdeterminant for tensors of larger dimension ($2 \times 2 \times 2 \times 2$, etc.). Our functions f_i are a different generalization of the construction, for which the format of the tensor is $n \times n \times n$ and does not yield hyperdeterminants.

Remark. For $n \geq 3$, f_i is not independent of the auxiliary indeterminates \mathbf{x} . In classical language, such a function might be called a *covariant* for the i th factor in $G(\mathbb{C})$, or a *concomitant* (see, for instance, [18, 24, 27]). Only in the case $n = 2$ is f_i an *invariant* in the strict sense of the term (i.e., associated to a one-dimensional representation). In section 5 we will see that f_i is associated with a higher-dimensional representation when $n > 2$.

We next use the function f_i to obtain a semialgebraic description of the orbit $\mathcal{D}(\mathbb{C})$. Since the vector \mathbf{x} has indeterminate entries, by a statement that $f_i(P; \mathbf{x}) = 0$ we mean that when f_i is evaluated at P , the resulting polynomial in \mathbf{x} is identically zero. Thus $f_i(P; \mathbf{x}) \neq 0$ means that at least one coefficient of an \mathbf{x} -monomial is nonzero at P .

THEOREM 4.1. *$P \in \mathcal{D}(\mathbb{C})$ if and only if for some $i \in \{1, 2, 3\}$, P satisfies the equations (3.1) and $f_i(P; \mathbf{x}) \neq 0$. Moreover, if these conditions hold for one $i \in \{1, 2, 3\}$, then they hold for all.*

Proof. Since $D *_i \mathbf{x} = \text{diag}(x_1, x_2, \dots, x_n)$, one computes

$$(4.7) \quad h_i(D; \mathbf{x}) = x_1 x_2 \cdots x_n,$$

so

$$f_i(D; \mathbf{x}) = (-1)^{n-1} \times \det \begin{pmatrix} 0 & x_3x_4x_5 \cdots x_n & x_2x_4x_5 \cdots x_n & \cdots & x_2x_3x_4 \cdots x_{n-1} \\ x_3x_4x_5 \cdots x_n & 0 & x_1x_4x_5 \cdots x_n & \cdots & x_1x_3x_4 \cdots x_{n-1} \\ \vdots & & \ddots & & \vdots \\ x_2x_3x_4 \cdots x_{n-1} & x_1x_3x_4 \cdots x_{n-1} & x_1x_2x_4 \cdots x_{n-1} & \cdots & 0 \end{pmatrix}$$

and

$$(x_1x_2 \cdots x_n)^2 f_i(D; \mathbf{x}) = (-1)^{n-1} \det \left((x_1x_2 \cdots x_n) \begin{pmatrix} 0 & 1 & 1 & \cdots & 1 \\ 1 & 0 & 1 & \cdots & 1 \\ \vdots & & \ddots & & \vdots \\ 1 & 1 & 1 & \cdots & 0 \end{pmatrix} \right);$$

hence

$$(4.8) \quad f_i(D; \mathbf{x}) = (n - 1)(x_1x_2 \cdots x_n)^{n-2}.$$

The transformation formula (4.6) then implies that if $P \in \mathcal{D}(\mathbb{C})$, then $f_i(P; \mathbf{x}) \neq 0$ for all i . That equations (3.1) hold when $P \in \mathcal{D}(\mathbb{C})$ is stated in Proposition 3.1.

Conversely, suppose for some i that $f_i(P; \mathbf{x}) \neq 0$ and equations (3.1) hold. Then P must be i -slice-nonsingular, since i -slice-singularity means that $h_i(P; \mathbf{x})$ is the zero polynomial, which implies that $f_i(P; \mathbf{x}) = 0$. Thus Proposition 3.3 applies, and we see that P is $G(\mathbb{C})$ -equivalent to a tensor P' with upper-triangular slices in index i . If such a P' had a slice with diagonal $\mathbf{0}$, then $\det(P' *_i \mathbf{x})$ would be independent of one of the x_i , so its Hessian would have a zero row (and column), implying that $f_i(P'; \mathbf{x})$ is the zero polynomial. By the transformation property (4.6), it would follow that $f_i(P; \mathbf{x}) = 0$ as well. Thus the slices of P' cannot have diagonals of $\mathbf{0}$, and Proposition 3.3 thus shows $P \in \mathcal{D}(\mathbb{C})$. \square

Note that the above theorem is concerned only with values of the f_i on the variety defined by equations (3.1); it makes no statement about f_i off this variety.

Since the variety defined by equations (3.1) is a supervariety of V_n , we immediately obtain the following.

COROLLARY 4.2. *$P \in \mathcal{D}(\mathbb{C})$ if and only if $P \in V_n$ and $f_i(P; \mathbf{x}) \neq 0$ for some (and hence all) $i \in \{1, 2, 3\}$.*

Equations (4.7) and (4.8) suggest consideration of the rational function

$$r_i(P; \mathbf{x}) = \frac{f_i(P; \mathbf{x})}{(n - 1)h_i(P; \mathbf{x})^{n-2}},$$

which is defined on the i -slice-nonsingular tensors and satisfies

$$(4.9) \quad r_i(D; \mathbf{x}) = 1.$$

Moreover, if $P' = P(g_1, g_2, g_3)$, then the transformation formulas (4.5) and (4.6) yield

$$(4.10) \quad r_i(P'; \mathbf{x}) = \det(g_1)^2 \det(g_2)^2 \det(g_3)^2 r_i(P, g_i \mathbf{x}).$$

Since $r_i(D; \mathbf{x})$ is independent of \mathbf{x} , (4.10) implies that $r_i(P; \mathbf{x})$ is independent of \mathbf{x} when $P \in \mathcal{D}(\mathbb{C})$ and thus, by continuity, when $P \in V_n$.

COROLLARY 4.3. $P \in \mathcal{D}(\mathbb{C})$ if and only if P satisfies the equations (3.1) and $r_i(P; \mathbf{x})$ is defined and nonzero for some (and hence all) $i \in \{1, 2, 3\}$.

In this statement the condition that P satisfies the equations (3.1) can, of course, be replaced by $P \in V_n$, as in Corollary 4.2.

While it is tempting to hope that $r_i(P; \mathbf{x})$ is independent of \mathbf{x} for all P , one can verify that this is not the case even when $n = 3$. Indeed, for the tensor

$$P = \left[\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right],$$

we have that

$$h_3(P; \mathbf{x}) = \det \begin{pmatrix} x & z & 0 \\ z & y & 0 \\ 0 & 0 & z \end{pmatrix} = xyz - z^3,$$

and

$$f_3(P; \mathbf{x}) = \det \begin{pmatrix} 0 & z & y \\ z & 0 & x \\ y & x & -6z \end{pmatrix} = 2xyz + 6z^3.$$

Thus $r_i(P; \mathbf{x})$ is neither independent of \mathbf{x} nor a polynomial function of \mathbf{x} . This example can easily be modified for $n \geq 3$.

Suppose now that one had a polynomial, $F(P)$, satisfying

$$(4.11) \quad F(P') = \det(g_1)^{2k} \det(g_2)^{2k} \det(g_3)^{2k} F(P),$$

when $P' = P(g_1, g_2, g_3)$. That is, suppose $F(P)$ is an invariant of weight $(2k, 2k, 2k)$ for $G(\mathbb{C})$. Then, provided that $F(D) \neq 0$, we may normalize so that $F(D) = 1$, and then observe that by their transformation formulas

$$(4.12) \quad r_i(P)^k = F(P) \text{ for } P \in \mathcal{D}(\mathbb{C}),$$

and thus, by continuity,

$$f_i(P; \mathbf{x})^k = (n - 1)^k h_i(P; \mathbf{x})^{k(n-2)} F(P)$$

for all $P \in V_n$. This yields the following.

PROPOSITION 4.4. *Let $F(P)$ be an invariant of weight $(2k, 2k, 2k)$ for $G(\mathbb{C})$ such that $F(D) \neq 0$. Then Theorem 4.1 and Corollary 4.2 remain true if f_i is replaced by the function*

$$G_i(P; \mathbf{x}) = h_i(P; \mathbf{x})^{k(n-2)} F(P).$$

While it is relatively straightforward to investigate the existence of polynomial invariants with weights of the type required for $F(P)$ for small n , it is less easy to give them explicitly. We discuss this further in the next section.

5. Representations and $n \times n \times n$ tensors. The previous section took an explicit, constructive approach to defining the invariants f_i . Here we turn to a more general understanding of the representation theory of $G(\mathbb{C})$. In particular, the transformation formula (4.6) of f_i and Proposition 4.4 indicate that studying all polynomials with good transformation properties under the group action might be useful. As a general background to the material in this section, we suggest [11, 14, 17, 25].

5.1. Representations and decompositions. As we will be concerned only with complex representations of complex groups, we suppress mention of \mathbb{C} in our notation in this section. We also let $V \cong \mathbb{C}^n$ (which should not be confused with our use of V_n for an algebraic variety in other sections).

Recall that a *representation* of a group G is a homomorphism $\rho : G \rightarrow GL(W)$. If W has no proper subspaces that are invariant under the action of G , then ρ is said to be *irreducible*. In particular, the irreducible representations of the general linear group on $V \cong \mathbb{C}^n$ are well understood to be

$$\rho_\lambda : GL(V) \cong GL(n, \mathbb{C}) \rightarrow GL(V^\lambda) \cong GL(n_\lambda, \mathbb{C}),$$

where the ρ_λ are labelled by integer partitions $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_\ell)$, with $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_\ell > 0$. We say λ is a partition of m and write $\lambda \vdash m$ and $|\lambda| = m$ when $\sum_{i=1}^\ell \lambda_i = m$. We refer to ℓ as the *depth* of λ . The representing space $V^\lambda \cong \mathbb{C}^{n_\lambda}$ (also referred to as a G -module), of dimension n_λ , can be expressed using the Schur functors $V^\lambda = \mathbb{S}_\lambda(V)$, but it is usually simpler to avoid explicitly doing so, and instead work with the characters of the representations. The characters $\{\lambda\} = \text{tr} \circ \rho_\lambda$, where “tr” denotes trace, are given by the Schur functions s_λ [24], with

$$\{\lambda\}(g) = s_\lambda(\xi),$$

where $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ are class parameters (eigenvalues) for g . Crucially, the Schur functions can be defined, and their properties explored, in a combinatorial manner quite independently of their role as characters for the general linear group [25].

The dimension $n_\lambda = \{\lambda\}(I)$ of the representation ρ_λ can be calculated by the hook length formula, which counts the number of semistandard tableaux of shape λ . For instance, the defining representation of $GL(V)$ is associated to the partition (1), with $n_{(1)} = n$, so $\{1\}$ denotes its trace and $s_{(1)}(\xi) = \xi_1 + \xi_2 + \dots + \xi_n$.

From two representations ρ_λ and $\rho_{\lambda'}$ of $GL(V)$, one can construct the tensor product representation $(\rho_\lambda \otimes \rho_{\lambda'})(g) := \rho_\lambda(g) \otimes \rho_{\lambda'}(g)$, with character denoted by $\{\lambda\} \otimes \{\lambda'\}$. While this representation may be reducible, its decomposition into irreducible representations of $GL(V)$ can be found using the pointwise product of Schur functions:

$$(\{\lambda\} \otimes \{\lambda'\})(g) = s_\lambda(\xi) s_{\lambda'}(\xi) = \sum_\alpha c_{\lambda\lambda'}^\alpha s_\alpha(\xi).$$

Here the multiplicities $c_{\lambda\lambda'}^\alpha$ are computable using the Littlewood–Richardson rule [25], with, for instance, software such as **Schur** [35].

The irreducible representations of $GL(n_1) \times GL(n_2) \times GL(n_3)$ are tensor products of the irreducible representations of the $GL(n_i)$:

$$\rho_{\lambda_1} \times \rho_{\lambda_2} \times \rho_{\lambda_3} : GL(n_1) \times GL(n_2) \times GL(n_3) \rightarrow GL(n_{\lambda_1}) \times GL(n_{\lambda_2}) \times GL(n_{\lambda_3}),$$

where $(\rho_{\lambda_1} \times \rho_{\lambda_2} \times \rho_{\lambda_3})(g_1, g_2, g_3) := \rho_{\lambda_1}(g_1) \otimes \rho_{\lambda_2}(g_2) \otimes \rho_{\lambda_3}(g_3)$. We denote the character of this representation by $\{\lambda_1\} \times \{\lambda_2\} \times \{\lambda_3\}$ to distinguish it from the product of characters of the sort described in the last paragraph.

The transformation of tensors $P \in U \cong V \otimes V \otimes V$ under elements $g = (g_1, g_2, g_3)$ of $G = G(\mathbb{C}) = GL(n, \mathbb{C}) \times GL(n, \mathbb{C}) \times GL(n, \mathbb{C})$ gives a representation of G with character $\{1\} \times \{1\} \times \{1\}$:

$$\text{tr}(g_1 \otimes g_2 \otimes g_3) = \text{tr}(g_1) \text{tr}(g_2) \text{tr}(g_3) = s_{(1)}(\xi) s_{(1)}(\xi') s_{(1)}(\xi'').$$

Denote the space of homogeneous polynomials of degree d in the components of tensors $P \in U$ by $\mathbb{C}[U]_d$. This space inherits an action of G by

$$f \mapsto g \circ f,$$

where $g \circ f(P) := f(P(g_1, g_2, g_3))$, and hence forms a G -module. By a standard argument, it is possible to identify $\mathbb{C}[U]_d \cong U^{(d)} = \mathbb{S}_{(d)}(U)$. While $\mathbb{C}[U]_d$ is usually not an irreducible G -module, through characters we can identify its decomposition into irreducible modules. This is done by applying the corresponding Schur function plethysm, denoted by $\underline{\otimes}$,

$$(5.1) \quad (\{1\} \times \{1\} \times \{1\}) \underline{\otimes} \{d\} = \sum_{\sigma_1, \sigma_2, \sigma_3 \vdash n} \gamma_{\sigma_1 \sigma_2 \sigma_3}^{(d)} \{\sigma_1\} \times \{\sigma_2\} \times \{\sigma_3\},$$

where the multiplicities $\gamma_{\sigma_1 \sigma_2 \sigma_3}^{(d)}$ can be calculated using standard group theory techniques implemented in software such as **Schur** [35], at least in cases of relatively small degree. (See below for an outline, and [32] for a more complete explanation.) In terms of G -modules, we can use (5.1) to identify

$$(5.2) \quad \mathbb{C}[U]_d \cong \bigoplus_{\sigma_1, \sigma_2, \sigma_3 \vdash n} \gamma_{\sigma_1 \sigma_2 \sigma_3}^{(d)} V^{\sigma_1} \otimes V^{\sigma_2} \otimes V^{\sigma_3}.$$

The primary focus of this section is to relate the functions $f_i(P; \mathbf{x})$ to this decomposition.

The coefficients in a plethysm formula such as (5.1) are the structure constants for Schur function “inner” products, denoted by $*$:

$$\{\alpha\} * \{\beta\} = \begin{cases} \sum_{\mu \vdash n} \gamma_{\alpha\beta}^{\mu} \{\mu\} & \text{if } |\alpha| = |\beta| = n, \\ 0 & \text{otherwise,} \end{cases}$$

where $\gamma_{\alpha\beta}^{\mu}$ is the multiplicity of the irreducible representation $\{\mu\}$ occurring in the decomposition of the tensor product of the irreducible representations $\{\alpha\}$ and $\{\beta\}$ in the symmetric group \mathfrak{S}_n . By linearity and associativity, we can similarly define

$$\{\alpha\} * \{\beta\} * \cdots * \{\zeta\} = \sum_{\mu \vdash n} \gamma_{\alpha\beta \dots \zeta}^{\mu} \{\mu\},$$

and the general plethysm of a product of defining representations is

$$(5.3) \quad (\{1\} \times \{1\} \times \cdots \times \{1\}) \underline{\otimes} \{\mu\} = \sum_{\alpha, \beta, \dots, \zeta \vdash |\mu|} \gamma_{\alpha\beta \dots \zeta}^{\mu} \{\alpha\} \times \{\beta\} \times \cdots \times \{\zeta\}.$$

Equation (5.1) is then the special case of a threefold product.

Remark. If in (5.3) the characters $\{1\}$ are replaced by $\{\rho\}, \{\sigma\}, \{\tau\}, \dots$, then the expansion on the right-hand side of (5.3) would be over the respective plethysms, $(\{\rho\} \underline{\otimes} \{\alpha\}), (\{\sigma\} \underline{\otimes} \{\beta\}), (\{\tau\} \underline{\otimes} \{\mu\}), \dots$. The simpler case of (5.3) arises since $\{1\} \underline{\otimes} \{\alpha\} \equiv \{\alpha\}$, $\{1\} \underline{\otimes} \{\beta\} \equiv \{\beta\}$, etc.

Remark. A familiar application of this theory is given by considering the $k \times k$ minors of an $n \times n$ matrix A . Under the action of $GL(n) \times GL(n) : A \mapsto g_1 A g_2^T$, for each integer $1 \leq k \leq n$, the span of the $k \times k$ minors of A is an invariant subspace of the homogeneous polynomials of degree k in the entries of A . In terms of Schur

function characters, the minors must therefore appear in the decomposition of the plethysm

$$(5.4) \quad (\{1\} \times \{1\}) \underline{\otimes} \{k\} = \sum_{\alpha, \beta \vdash k} \gamma_{\alpha\beta}^{(k)} \{\alpha\} \times \{\beta\}.$$

Adopting the standard shorthand of using exponents to signify repeated integers in a partition, (k) and $(1^k) = (1, 1, \dots, 1)$ label the one-dimensional “trivial” and “sign” representations of the symmetric group \mathfrak{S}_k , respectively. The calculation $\gamma_{(1^k)(1^k)}^{(k)} = 1$ follows immediately from the tensor product $sgn(\sigma) \otimes sgn(\sigma) = sgn(\sigma)sgn(\sigma) = 1$ for all $\sigma \in \mathfrak{S}_k$.

Thus the character $\{1^k\} \times \{1^k\}$ appears exactly once in the above expansion. Via the hook length formula one finds $\dim(\rho_{(1^k)} \times \rho_{(1^k)}) = \dim(\rho_{(1^k)})^2 = \binom{n}{k}^2$, and the associated irreducible subspace is confirmed to be that spanned by the $k \times k$ minors of A .

5.2. One-dimensional representations. As first use of this viewpoint, we investigate polynomial invariants of $G(\mathbb{C})$, that is, one-dimensional representations within the vector space of polynomial functions. Proposition 4.4 indicates that any polynomial F satisfying (4.11) for any positive integer k can be used in characterizing the points on $\mathcal{D}(\mathbb{C})$ (provided that $F(D) \neq 0$), so we seek to find such F .

The one-dimensional representation $\det(g)^k$ of $GL(n)$ has character $\{k^n\}$. Via calculations with Schur [35], by decomposing $\mathbb{C}[U]$ for $n = 2, 3, 4$, we find that there are polynomials in the entries of P transforming as $\{2^2\} \times \{2^2\} \times \{2^2\}$, $\{2^3\} \times \{2^3\} \times \{2^3\}$, $\{2^4\} \times \{2^4\} \times \{2^4\}$ of degree $d = 4, 6, 8$ for $n = 2, 3, 4$, respectively. Since the multiplicities of the representation in the decomposition are found to be

$$1 = \gamma_{(2^2)(2^2)(2^2)}^{(4)} = \gamma_{(2^3)(2^3)(2^3)}^{(6)} = \gamma_{(2^4)(2^4)(2^4)}^{(8)},$$

the polynomial is uniquely determined up to scaling. Thus for $n = 2, 3, 4$ we denote this function by τ_n and call it the n -tangle, since Cayley’s hyperdeterminant $\tau_2 = \Delta$ is called the tangle in the physics literature, and τ_2, τ_3 , and τ_4 all transform with weight $(2, 2, 2)$. (We have not yet fixed a choice of scaling for τ_3, τ_4 , but we will below.) This progression stops with $n = 4$, but for $n = 5, 6$ there are invariants transforming with weight $(3, 3, 3)$, which again occur with multiplicity 1 and so are unique up to scaling. These results are summarized in Table 5.1. Beyond $n = 7$, computations with Schur become prohibitive, although we verified that there are no weight $(2, 2, 2)$ representations for $8 \leq n \leq 16$.

TABLE 5.1

Multiplicities $\gamma_{(k^n)(k^n)(k^n)}^{(nk)}$ of one-dimensional representations of weight (k, k, k) in the space of homogeneous polynomials of degree nk in the entries of tensors of format $n \times n \times n$, as computed by Schur.

Tensor format	Weight		
	(2,2,2)	(3,3,3)	(4,4,4)
$2 \times 2 \times 2$	1	0	1
$3 \times 3 \times 3$	1	1	2
$4 \times 4 \times 4$	1	1	5
$5 \times 5 \times 5$	0	1	6
$6 \times 6 \times 6$	0	1	
$7 \times 7 \times 7$	0	0	

While the existence of invariants with the desired transformation property for $n < 6$ is now established, to use them in Proposition 4.4 requires that we also know that they do not vanish at D . Just as this can be easily seen from the explicit formula (1.1) for the tangle when $n = 2$, to establish this in the cases $n = 3, 4$, we turn to an explicit construction of the invariant.

For $n = 3$, let ϵ_{ijk} be the antisymmetric Levi–Civita tensor with $\epsilon_{123} = \epsilon_{231} = \epsilon_{321} = 1$, $\epsilon_{213} = \epsilon_{312} = \epsilon_{132} = -1$, and $\epsilon_{ijk} = 0$ otherwise. Consider the degree 6 polynomial $\tau_3 : U \rightarrow \mathbb{C}$ defined by

$$\tau_3(P) = \sum_1^3 P_{i_1 i_2 i_3} P_{j_1 j_2 j_3} P_{k_1 k_2 k_3} P_{l_1 l_2 l_3} P_{m_1 m_2 m_3} P_{n_1 n_2 n_3} \times \epsilon_{i_1 j_1 k_1} \epsilon_{j_2 k_2 l_2} \epsilon_{k_3 l_3 m_3} \epsilon_{l_1 m_1 n_1} \epsilon_{m_2 n_2 i_2} \epsilon_{n_3 i_3 j_3},$$

where all 18 indices run from 1 to 3 in the sum. Since ϵ_{ijk} defines a one-dimensional representation of $GL(3)$ by

$$\sum_{1 \leq i', j', k' \leq 3} g(i, i') g(j, j') g(k, k') \epsilon_{i' j' k'} = \det(g) \epsilon_{ijk} \quad \text{for } g \in GL(3, \mathbb{C}),$$

it is straightforward to check that

$$\tau_3(P(g_1, g_2, g_3)) = \det(g_1)^2 \det(g_2)^2 \det(g_3)^2 \tau_3(P)$$

for all $(g_1, g_2, g_3) \in G(\mathbb{C})$.

As noted in [33], expanding τ_3 as a polynomial yields 1152 terms [2], and thus it is not the zero polynomial. But we wish to establish the stronger statement that $\tau_3(D) \neq 0$. Expressing D in components $D_{ijk} = \delta_{ij} \delta_{jk}$, one first finds

$$\tau_3(D) = \sum_{1 \leq i, j, k, l, m, n \leq 3} \epsilon_{ijk} \epsilon_{jkl} \epsilon_{klm} \epsilon_{lmn} \epsilon_{mni} \epsilon_{nij}.$$

Now the first factor in the summand, ϵ_{ijk} , is zero unless $i, j, k \in \{1, 2, 3\}$ are distinct. Then the product of the first two factors is zero unless additionally $l = i$. Considering the remaining four ϵ factors in this way, nonzero contributions also require $m = j$ and $n = k$. Thus

$$\tau_3(D) = \sum_{i, j, k \text{ distinct}} \epsilon_{ijk}^2 \epsilon_{jki}^2 \epsilon_{kij}^2 = 6.$$

This confirms both that τ_3 is a nonzero polynomial and that it evaluates to a positive value on the diagonal tensor.

Similar considerations give the 4-tangle. With ϵ_{ijkl} denoting the sign of the permutation $(ijkl)$ when i, j, k, l are distinct, and 0 otherwise, let

$$\tau_4(P) = \sum_1^4 P_{i_1 i_2 i_3} P_{j_1 j_2 j_3} P_{k_1 k_2 k_3} P_{l_1 l_2 l_3} P_{m_1 m_2 m_3} P_{n_1 n_2 n_3} P_{r_1 r_2 r_3} P_{s_1 s_2 s_3} \times \epsilon_{i_1 j_1 k_1 l_1} \epsilon_{m_1 n_1 r_1 s_1} \epsilon_{i_2 l_2 m_2 s_2} \epsilon_{j_2 k_2 n_2 r_2} \epsilon_{i_3 j_3 m_3 n_3} \epsilon_{k_3 l_3 r_3 s_3},$$

where all 24 indices run from 1 to 4 in the sum. A polynomial expansion of this has 431,424 terms [2]. By an argument analogous to that for τ_3 , one sees that the only nonzero terms in

$$\tau_4(D) = \sum_1^4 \epsilon_{ijkl} \epsilon_{mnr s} \epsilon_{ilms} \epsilon_{jknr} \epsilon_{ijmn} \epsilon_{klrs}$$

occur when $m = k, n = l, r = i, s = j$, and thus that $\tau_4(D) = 24$, confirming that $\tau_4(D)$ is also nonzero.

We do not have an explicit construction of invariants for $n = 5, 6$ that have weight $(3, 3, 3)$, and we therefore do not know whether they vanish at D .

5.3. Higher-dimensional representations. The functions f_i constructed in section 4 are not invariants when $n > 2$, due to their dependence on the auxiliary variables \mathbf{x} , and thus do not define one-dimensional representations of $G(\mathbb{C})$. We therefore turn to studying higher-dimensional representations in polynomials.

A consequence of considering irreducible modules of polynomials is that statements concerning polynomials vanishing on $G(\mathbb{C})$ -invariant sets which apply to a specific element of the module must also apply to the module as a whole. This is formalized in the following two lemmas.

LEMMA 5.1. *Let $f \in \mathbb{C}[U]_d$. Then $\langle \{g \circ f : g \in G\} \rangle_{\mathbb{C}}$, the linear span of the G -orbit of f , is a G -module. In particular, if W is an irreducible submodule of $\mathbb{C}[U]_d$, and $0 \neq f \in W$, then $\langle \{g \circ f : g \in G\} \rangle_{\mathbb{C}} = W$.*

Proof. If $p \in \langle \{g \circ f : g \in G\} \rangle_{\mathbb{C}}$, then, for some finite subset $S \subset G$ and $c_h \in \mathbb{C}$,

$$p = \sum_{h \in S} c_h (h \circ f).$$

Thus if $g \in G$,

$$g \circ p = \sum_{h \in S} c_h (gh \circ f) \in \langle \{g \circ f : g \in G\} \rangle_{\mathbb{C}},$$

so the linear span of the orbit is a G -module.

Now for any G -module W and $f \in W$, $\langle \{g \circ f : g \in G\} \rangle_{\mathbb{C}} \subseteq W$. Irreducibility of W and $f \neq 0$ thus implies that $\langle \{g \circ f : g \in G\} \rangle_{\mathbb{C}} = W$. \square

Although as stated here this lemma applies to f in a G -module of polynomials, the result is a standard one for any G -module. Though also stated for polynomials, the next result holds more generally for G -modules of functions where the action of G arises from an action on their domain.

LEMMA 5.2. *Let \mathcal{S} be a G -invariant subset of U , and let $W \subseteq \mathbb{C}[U]_d$ be an irreducible G -module. Then $f|_{\mathcal{S}} \equiv 0$ for some nonzero $f \in W$ if and only if $p|_{\mathcal{S}} \equiv 0$ for all $p \in W$.*

Proof. Let $f \in W$ with $f|_{\mathcal{S}} \equiv 0$. Then for any $p \in W, P \in U$, by the preceding lemma

$$p(P) = \sum_{h \in S} c_h (h \circ f)(P) = \sum_{h \in S} c_h f(Ph).$$

But $P \in \mathcal{S}$ implies $Ph \in \mathcal{S}$, so this shows $p|_{\mathcal{S}} \equiv 0$. \square

There are several G -invariant sets of interest in this paper. They are $\mathcal{D}(\mathbb{C})$, the orbit of the tensor D ; $V_n = \overline{\mathcal{D}(\mathbb{C})}$, the orbit closure; and $V_n \setminus \mathcal{D}(\mathbb{C})$, the complement of the orbit in its closure. However, by continuity, polynomials that vanish on $\mathcal{D}(\mathbb{C})$ vanish on its closure V_n as well, so investigating polynomials vanishing on either of these sets leads to polynomials in the defining ideal of the variety V_n . Indeed, some such polynomials are given in Proposition 3.1, though not from the point of view of representations.

The functions $f_i(P; \mathbf{x})$ constructed in section 4, however, are zero on $V_n \setminus \mathcal{D}(\mathbb{C})$, and nonzero on $\mathcal{D}(\mathbb{C})$, by Corollary 4.2. Thus Lemma 5.2 suggests relating the classical

viewpoint on f_i used in their construction to the language of representations. Without loss of generality we focus on $f_3(P; \mathbf{x})$ and think of $f_3(P; \mathbf{x})$ as providing a set of functions $f_3(\cdot; \mathbf{x}) : P \mapsto f_3(P; \mathbf{x})$ parametrized by the auxiliary variables \mathbf{x} .

Notationally, we use nonnegative integer vectors $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)$ to express a monomial $x^\alpha := x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n}$ of total degree $d = \sum_{1 \leq i \leq n} \alpha_i$. Then

$$f_3(P; \mathbf{x}) = \sum_{\alpha} p_{\alpha}(P) x^{\alpha},$$

with coefficients $p_{\alpha} \in \mathbb{C}[U]_{n^2}$, where the monomials x^{α} are interpreted as basis elements for $\mathbb{C}[V]_d \cong V^{(d)}$, with $d = n(n - 2)$.

One can see directly that no p_{α} is identically zero: To start, note that (4.8) gives

$$\begin{aligned} (5.5) \quad \sum_{\alpha} p_{\alpha}(D(I, I, g_3)) x^{\alpha} &= f_3(D(I, I, g_3); \mathbf{x}) \\ &= \det(g_3)^2 f_3(D; g_3 \mathbf{x}) \\ &= \det(g_3)^2 (n - 1) ((g_3 \mathbf{x})_1 (g_3 \mathbf{x})_2 \cdots (g_3 \mathbf{x})_n)^{n-2}. \end{aligned}$$

Choosing $g_3 \in GL(n, \mathbb{C})$ with strictly positive entries, every possible monomial x^{α} appears in the expansion of $((g_3 \mathbf{x})_1 (g_3 \mathbf{x})_2 \cdots (g_3 \mathbf{x})_n)^{n-2}$. Hence $p_{\alpha}(D(I, I, g_3)) \neq 0$ for all α .

To understand the transformation of p_{α} under $g = (g_1, g_2, g_3) \in G(\mathbb{C})$, observe that g_3 maps the monomial x^{α} of total degree d by

$$\begin{aligned} x^{\alpha} &\mapsto \left(\sum_{i_1=1}^n g_3(1, i_1) x_{i_1} \right)^{\alpha_1} \left(\sum_{i_2=1}^n g_3(2, i_2) x_{i_2} \right)^{\alpha_2} \cdots \left(\sum_{i_n=1}^n g_3(n, i_n) x_{i_n} \right)^{\alpha_n} \\ &:= \sum_{\beta} \tilde{g}_3(\alpha, \beta) x^{\beta}. \end{aligned}$$

The matrix elements $\tilde{g}_3(\alpha, \beta)$ provide precisely the irreducible representation $\rho_{\lambda} : GL(n, \mathbb{C}) \rightarrow GL(n_{\lambda}, \mathbb{C})$ with $\lambda = (d)$ and $\tilde{g}_3 = \rho_{(d)}(g_3)$. The polynomials $p_{\alpha}(P)$ therefore transform under $G(\mathbb{C})$ as

$$(5.6) \quad p_{\alpha} \mapsto \det(g_1)^n \det(g_2)^n \det(g_3)^2 \sum_{\beta} p_{\beta} \tilde{g}_3(\beta, \alpha).$$

This formula also implies that the p_{α} are independent: If $\mathbf{c} = (c_{\alpha})$ specifies a dependency relation $\sum c_{\alpha} p_{\alpha} = 0$, then from (5.6) it follows that $\mathbf{d} = \tilde{g}_3 \mathbf{c}$ gives another dependency relation for every choice of g_3 . By the irreducibility of $\rho_{(d)}$, and varying g_3 , this can happen only if all p_{α} vanish identically, which they do not.

Now let W be the span of $\{p_{\alpha}\}$. Since the p_{α} are independent, (5.6) defines a linear map on W , making W an irreducible $G(\mathbb{C})$ -module. The character of the corresponding representation of $G(\mathbb{C})$ is the product $\{n^n\} \times \{n^n\} \times (\{2^n\} \otimes \{d\})$, with $d = n(n - 2)$. Application of the Littlewood–Richardson rule [25] shows that, as a character of $GL(n, \mathbb{C})$ where partitions of depth greater than n are excluded, the third factor is $\{2^n\} \otimes \{d\} = \{2 + n(n - 2), 2^{n-1}\}$.

Thus the polynomial $f_3(P; \mathbf{x})$ is associated to a module W of polynomials in the entries of P alone that transforms as $\{n^n\} \times \{n^n\} \times \{2 + n(n - 2), 2^{n-1}\}$. The

dimension of such an irreducible module is calculated by the hook length formula as

$$\begin{aligned} \dim((n^n)) \times \dim((n^n)) \times \dim((2 + n(n - 2), 2^{n-1})) &= 1 \times 1 \times \binom{n(n - 2) + (n - 1)}{n(n - 2)} \\ &= \binom{n^2 - n - 1}{n - 1}. \end{aligned}$$

This result is not a surprise, since it is the dimension of the space of homogeneous polynomials in n variables of degree $n(n - 2)$, i.e., the cardinality of the basis $\{p_\alpha\}$ of W .

The multiplicity of modules transforming as $\{n^n\} \times \{n^n\} \times \{2 + n(n - 2), 2^{n-1}\}$ in the decomposition of $\mathbb{C}[U]_{n^2}$ can be calculated, at least for a few small values of n , using Schur, and are given in Table 5.2. Note that in the notation for the multiplicity

$$\gamma_{(n^n)(n^n)(2+n(n-2),2^{(n-1)})}^{(n^2)},$$

whose values are given in the table, the superscript (n^2) should be read as a one-part partition of the integer n^2 , while the subscripts (n^n) denote the n -part partition (n, n, \dots, n) of n^2 in the standard shorthand notation for partitions.

TABLE 5.2

Irreducible modules associated to the functions f_3 on $n \times n \times n$ tensors, along with their multiplicities $\gamma_{(n^n)(n^n)(2+n(n-2),2^{(n-1)})}^{(n^2)}$ and dimensions $\binom{n^2-n-1}{n-1}$, in the space of homogeneous polynomials of degree n^2 in the entries of P .

n	Module	Multiplicity	Dimension
2	$\{2^2\} \times \{2^2\} \times \{2^2\}$	1	1
3	$\{3^3\} \times \{3^3\} \times \{5, 2^2\}$	2	10
4	$\{4^4\} \times \{4^4\} \times \{10, 2^3\}$	5	165
5	$\{5^5\} \times \{5^5\} \times \{17, 2^4\}$	10	3876

For $n = 3$, the multiplicity of 2 shown in Table 5.2 allows not only for the existence of f_i but also for an additional function transforming in the same way (or, more accurately, for a two-dimensional space of such functions containing f_i). Indeed, from the transformation formula (4.5) for h_i and the fact that τ_3 is an invariant of weight $(2, 2, 2)$, the function $G_i(P; \mathbf{x}) = h_i(P; \mathbf{x})\tau_3(P)$ will be such a function. This is precisely the construction given in Proposition 4.4, and since $\tau_3(D) \neq 0$, that result applies. By the discussion preceding that proposition, f_i/h_i is not independent of \mathbf{x} and thus cannot be a multiple of τ_3 . Thus f_i and G_i are independent.

Similarly, for $n = 4$, $G_i = h_i^2\tau_4$ and f_i transform in the same way and can be seen to be independent. While Table 5.2 indicates the existence of three other independent functions with the same transformation property, we have no explicit understanding of them and do not know what, if anything, they (or linear combinations of them) indicate about tensor rank.

6. Application to real tensors. In this section we investigate real tensors in $\mathcal{D}(\mathbb{C})$. As Corollary 4.2 gives a semialgebraic description of $\mathcal{D}(\mathbb{C})$, it is natural to seek a similar description of $\mathcal{D}(\mathbb{R})$. To obtain conditions that define $\mathcal{D}(\mathbb{R})$ we should, of course, include the additional condition that a tensor P be real. However, even in the case that $n = 2$, this is not sufficient to define $\mathcal{D}(\mathbb{R})$; one also needs that $\Delta(P) > 0$ [10].

Using our functions f_i as a tool, our main results are as follows. For $n = 3$ the sign of an invariant function can be used to distinguish $\mathcal{D}(\mathbb{R})$, extending the $n = 2$ result in this single case. For all $n \geq 2$, the zero set of our f_i partitions the real points in $\mathcal{D}(\mathbb{C})$ into connected components. Within a component, all tensors have the same number of complex conjugate pairs of rank-1 tensors in their rank decompositions. Moreover, with one exception, there is a single component for each allowable number of pairs. In particular one component of the real points is $\mathcal{D}(\mathbb{R})$.

Let $V_n(\mathbb{R}) = V_n \cap \mathbb{R}^{n^3}$ denote the real points on V_n .

LEMMA 6.1. *Suppose $P \in \mathcal{D}(\mathbb{C}) \cap V_n(\mathbb{R})$. Then, up to simultaneous permutation of the rows of the g_i , P can be uniquely expressed as $P = D(g_1, g_2, g_3)$, subject to the following conditions:*

- (i) *The first nonzero entry of every row of g_1 and g_2 is 1.*
- (ii) *For some $k \leq n/2$, the first $2k$ rows of every g_i are complex (and neither real nor purely imaginary) in conjugate pairs, and the remaining rows are real. Thus P has a unique decomposition into complex rank-1 components with $2k$ complex components in conjugate pairs and $n - 2k$ real components.*

Proof. For some $g_1, g_2, g_3 \in GL(n, \mathbb{C})$, $P = D(g_1, g_2, g_3)$, which implies

$$(6.1) \quad P = \sum_{i=1}^n \mathbf{g}_1^i \otimes \mathbf{g}_2^i \otimes \mathbf{g}_3^i,$$

where \mathbf{g}_j^i is the i th row of g_j . Since the rows of each g_j are independent, Kruskal’s theorem [20, 21, 29] implies that this rank-1 decomposition is unique, up to ordering of the summands. However the individual vectors \mathbf{g}_j^i can be multiplied by scalars a_j^i , as long as $a_1^i a_2^i a_3^i = 1$. Requiring the first nonzero entries in each row of g_1, g_2 to be 1 removes that freedom.

Since P is real, the complex conjugate of the decomposition in (6.1) must give the same decomposition, up to order of the summands. Thus for each i either $\mathbf{g}_1^i \otimes \mathbf{g}_2^i \otimes \mathbf{g}_3^i$ is real or its complex conjugate also appears as a summand.

We may thus simultaneously permute the rows of the g_j , so for $i = 1, \dots, k$

$$\mathbf{g}_1^{2i-1} \otimes \mathbf{g}_2^{2i-1} \otimes \mathbf{g}_3^{2i-1} = \overline{\mathbf{g}_1^{2i}} \otimes \overline{\mathbf{g}_2^{2i}} \otimes \overline{\mathbf{g}_3^{2i}},$$

and for $2k < i \leq n$ the summand is real.

Having done this, since each $\mathbf{g}_1^i, \mathbf{g}_2^i$ has an entry of 1, from the conjugate summands we first see that $\mathbf{g}_3^{2i-1} = \overline{\mathbf{g}_3^{2i}}$ for $i = 1, \dots, k$ and then that similar statements hold for the rows of g_1 and g_2 . Thus all three g_i have the first $2k$ rows in conjugate pairs. Moreover, none of these first $2k$ rows of any g_i is real, lest there be a repeated row, contradicting that $g_i \in GL(n, \mathbb{C})$. Likewise, these rows are not purely imaginary.

An analogous argument shows that the remaining rows of the g_i are real. \square

For $P \in \mathcal{D}(\mathbb{C}) \cap V_n(\mathbb{R})$, we refer to the ordered pair $(n - 2k, k)$ of this lemma as the *signature* of P . For $P \notin \mathcal{D}(\mathbb{C}) \cap V_n(\mathbb{R})$, the rank-1 tensor decomposition may not be unique, so we leave the signature undefined. The orbit $\mathcal{D}(\mathbb{R})$ thus comprises those $P \in \mathcal{D}(\mathbb{C})$ with signature $(n, 0)$.

Information on the signature of a tensor can be obtained from the value of the function r_i defined in section 4, as we now show.

THEOREM 6.2. *For all n , the set of tensors $P \in V_n(\mathbb{R})$ for which $r_i(P) > 0$ is precisely those tensors of signature $(n - 2k, k)$ with k even.*

In particular, for $n = 2$, $\mathcal{D}(\mathbb{R})$ is precisely the set of tensors $P \in V_2(\mathbb{R})$ with $\Delta(P) > 0$, and for $n = 3$, $\mathcal{D}(\mathbb{R})$ is precisely the set of tensors $P \in V_3(\mathbb{R})$ with $\tau_3(P) > 0$.

For any $n \geq 3$, $\mathcal{D}(\mathbb{R})$ is precisely the set of tensors $P \in V_n(\mathbb{R})$ with $f_i(P; \mathbf{x}) \neq 0$ such that when $f_i(P; \mathbf{x})$ is factored into linear forms as

$$f_i(P; \mathbf{x}) = c \prod_{j=1}^n l_j(\mathbf{x})^{n-2},$$

then all of the linear forms l_j may be taken to be real.

In this statement the conditions $P \in V_n(\mathbb{R})$ with $f_i(P; \mathbf{x}) \neq 0$ can be replaced, by Theorem 4.1 and Corollary 4.2, by an explicit collection of polynomial equalities and $f_i(P; \mathbf{x}) \neq 0$.

Proof. If P has signature $(n - 2k, k)$, then $P = D(g_1, g_2, g_3)$, where each g_i has exactly k pairs of complex conjugate rows. Thus $\bar{g}_i = \sigma g_i$ for some permutation σ with $\det \sigma = (-1)^k$. This implies that $\overline{\det g_i} = (-1)^k \det g_i$, so $\det g_i$ is real or purely imaginary according to whether k is even or odd. Thus $(\det g_i)^2$ is positive or negative according to whether k is even or odd. Using the properties of r_i given in formulas (4.10) and (4.9), the first claim is established.

Note that for $n = 2, 3$, the only allowable even value for k is zero. The claim for $n = 2$ is then immediate since $r_i = \Delta$ in that case. For $n = 3$ the claim follows similarly, from the facts that τ_3 is of weight $(2, 2, 2)$ and $\tau_3(D) > 0$, so, by equation (4.12), τ_3 is a positive multiple of r_i when restricted to $\mathcal{D}(\mathbb{C})$.

For arbitrary $n \geq 3$, $P \in V_n$, and $f_i(P; \mathbf{x}) \neq 0$ is equivalent to $P \in \mathcal{D}(\mathbb{C})$ by Corollary 4.2. By (4.6) and (4.8), we have that for $P = D(g_1, g_2, g_3)$,

$$f_i(P; \mathbf{x}) = c' f_i(D, g_i; \mathbf{x}) = c \prod_{j=1}^n l_j(\mathbf{x})^{n-2}$$

for some scalars c', c and linear forms l_j defined by the rows of g_i . If $P \in \mathcal{D}(\mathbb{R})$, so that the g_i may be taken to be real, then $f_i(P; \mathbf{x})$ has a factorization using real linear forms. For $n \geq 3$ and $P \notin \mathcal{D}(\mathbb{R})$, by Lemma 6.1 such a factorization exists with at least two l_j complex and not multiples of real forms. By unique factorization in the ring $\mathbb{C}[\mathbf{x}]$, there can be no factorization into powers of real linear forms in this case. \square

Note that the statement in this theorem about the factorization into linear forms of $f_i(P; \mathbf{x})$ could be replaced with a similar one about $h_i(P; \mathbf{x})$.

We next consider the connected components obtained from $V_n(\mathbb{R})$ by removing the zero set of an f_i . Let $Z_n = \{P \in V_n(\mathbb{R}) \mid f_i(P; \mathbf{x}) = 0\}$, and note that Z_n is independent of the choice of $i \in \{1, 2, 3\}$, by Corollary 4.2.

THEOREM 6.3. *On each path component of $V_n(\mathbb{R}) \setminus Z_n$ the signature is constant. For each $0 \leq k < n/2$, there is one path component of signature $(n - 2k, k)$. When n is even, there are four components with signature $(0, n/2)$.*

Proof. Suppose a component of $V_n(\mathbb{R}) \setminus Z_n$ contains tensors of signature $(n - 2k, k)$ for two different values of k . Let $k(P)$ denote the second term in the signature $(n - 2k, k)$ of a tensor P . Then in this component we can choose a tensor P_0 and a sequence of tensors P_1, P_2, P_3, \dots with $\lim_{\ell \rightarrow \infty} P_\ell = P_0$, $k(P_0) = k_0 \neq k_1 = k(P_\ell)$ for $\ell \geq 1$. But then

$$(6.2) \quad \lim_{\ell \rightarrow \infty} f_i(P_\ell; \mathbf{x}) = f_i(P_0; \mathbf{x}).$$

By Theorem 6.2, for each $\ell > 0$ the function $f_i(P_\ell; \mathbf{x})$ factors as

$$(6.3) \quad f_i(P_\ell; \mathbf{x}) = c_\ell \prod_{j=1}^n l_{\ell,j}(\mathbf{x})^{n-2},$$

where we assume that the linear forms in this factorization have been normalized so that

$$l_{\ell,j}(\mathbf{x}) = \mathbf{u}_{\ell,j} \cdot \mathbf{x},$$

with $\|\mathbf{u}_{\ell,j}\| = 1$ for all j . We further assume that complex vectors $\mathbf{u}_{\ell,2j-1} = \overline{\mathbf{u}_{\ell,2j}}$ for $1 \leq j \leq k_1$ are associated to the nonreal linear forms, and real vectors $\mathbf{u}_{\ell,j}$ $2k_1 < j \leq n$ to the real ones.

By compactness of the unit sphere in \mathbb{C}^n , passing to a subsequence of $\{P_\ell\}$, we may assume that for each j

$$\lim_{\ell \rightarrow \infty} \mathbf{u}_{\ell,j} = \mathbf{u}_j$$

for some unit vector \mathbf{u}_j . Let $l_j(\mathbf{x}) = \mathbf{u}_j \cdot \mathbf{x}$. Now by (6.2) and (6.3) we see that

$$f_i(P_0; \mathbf{x}) = \left(\lim_{\ell \rightarrow \infty} c_\ell \right) \prod_{j=1}^n l_j(\mathbf{x})^{n-2}.$$

Since for $n - 2k_1$ values of j the $\mathbf{u}_{\ell,j}$ are real, at least this many of the \mathbf{u}_j are. Thus, since $k_0 \neq k_1$, we have $k_0 < k_1$.

However, $k_0 < k_1$ implies that for some $1 \leq j \leq k_1$, $\mathbf{u}_{2j} = \lim_{\ell \rightarrow \infty} \mathbf{u}_{\ell,2j}$ is real. But since $\mathbf{u}_{\ell,2j-1} = \overline{\mathbf{u}_{\ell,2j}}$, this means that $\mathbf{u}_{2j} = \mathbf{u}_{2j-1}$. This is impossible, as these vectors are, up to scaling, rows of some g_i where $P_0 = D(g_1, g_2, g_3)$, and thus must be independent.

Thus the signature is constant on each component.

The number of path components that exist for any fixed value of k , $0 \leq k \leq n/2$, is always at least one, since one can construct a real tensor of signature $(n - 2k, k)$. We now turn to giving an upper bound on the number of such components.

Note first that a $2 \times n$ matrix with complex conjugate rows can be expressed as

$$\begin{pmatrix} 1 & i \\ 1 & -i \end{pmatrix} \begin{pmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \end{pmatrix}$$

for row vectors $\mathbf{r}_1, \mathbf{r}_2 \in \mathbb{R}^n$. Thus if J_k is an $n \times n$ block diagonal matrix with k blocks of $\begin{pmatrix} 1 & i \\ 1 & -i \end{pmatrix}$ and $n - 2k$ singleton blocks of 1, then tensors in $V_n(\mathbb{R}) \setminus Z_n$ with signature $(n - 2k, k)$ form the $G(\mathbb{R})$ -orbit of $D(J_k, J_k, J_k)$. We thus seek to bound the number of path components of this orbit.

Recall that $GL_n(\mathbb{R})$ has two path components, $GL_n^+(\mathbb{R})$ and $GL_n^-(\mathbb{R})$, with membership according to the sign of the determinant. Then $G(\mathbb{R})$ has eight path components, one of which is

$$G^+(\mathbb{R}) = GL_n^+(\mathbb{R}) \times GL_n^+(\mathbb{R}) \times GL_n^+(\mathbb{R}).$$

The trivial bound on the number of components of the $G(\mathbb{R})$ orbit of $D(J_k, J_k, J_k)$ is thus 8.

Suppose $k \leq (n - 1)/2$, so tensors of signature $(n - 2k, k)$ have at least one real rank-1 component, and J_k has at least one 1×1 diagonal block, which we assume is the last. Let $K = \text{diag}(1, 1, \dots, 1, -1)$, and observe that since $J_k K = K J_k$,

$$\begin{aligned} D(J_k, J_k, J_k)(K, I, I) &= D(K, I, I)(J_k, J_k, J_k) \\ &= D(I, K, I)(J_k, J_k, J_k) \\ &= D(J_k, J_k, J_k)(I, K, I), \end{aligned}$$

and similarly $D(J_k, J_k, J_k)(I, K, I) = D(J_k, J_k, J_k)(I, I, K)$. Since $\det(K) = -1$, this implies that the $G(\mathbb{R})$ -orbit of $D(J_k, J_k, J_k)$ is the union of the $G^+(\mathbb{R})$ -orbits of $D(J_k, J_k, J_k)$ and $D(J_k, J_k, J_k)(K, I, I)$. To show that there is only one path component, we now need only show that $D(J_k, J_k, J_k)$ and $D(J_k, J_k, J_k)(K, I, I)$ are in the same component.

If $k < (n - 1)/2$, then J_k has at least two 1×1 diagonal blocks in the last positions. Let

$$R_2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Then one checks that

$$D_2(\sigma_2, R_2, R_2) = D_2.$$

Letting R be the $n \times n$ block diagonal matrix $R = \text{diag}(1, 1, \dots, 1, R_2)$, and letting $\sigma = \text{diag}(1, 1, \dots, 1, \sigma_2)$, it follows that

$$D(J_k, J_k, J_k)(\sigma, R, R) = D(J_k, J_k, J_k).$$

Since $\det(R) = 1$ and $\det(\sigma) = -1 = \det(K)$, $D(J_k, J_k, J_k)(\sigma, R, R)$ is in the $G^+(\mathbb{R})$ -orbit, and hence the path component, of $D(J_k, J_k, J_k)(K, I, I)$. Thus $D(J_k, J_k, J_k)$ and $D(J_k, J_k, J_k)(K, I, I)$ are in the same component, and there is only one path component of signature $(n - 2k, k)$.

In the case when n is odd and $k = (n - 1)/2$, J_k has a single 1×1 block in the last position. Observe that

$$\begin{pmatrix} 1 & i \\ 1 & -i \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & i \\ 1 & -i \end{pmatrix},$$

so if $L = \text{diag}(1, 1, \dots, 1, -1, 1)$, then

$$J_k L = \text{diag}(1, 1, \dots, 1, \sigma_2, 1) J_k.$$

Thus

$$D(J_k, J_k, J_k)(L, L, L) = D(J_k, J_k, J_k).$$

Now $D(J_k, J_k, J_k)(LK, LK, LK)$ is in the $G^+(\mathbb{R})$ -orbit, and thus path component, of $D(J_k, J_k, J_k)$, but

$$\begin{aligned} D(J_k, J_k, J_k)(LK, LK, LK) &= D(J_k, J_k, J_k)(L, L, L)(K, K, K) \\ &= D(J_k, J_k, J_k)(K, K, K) \\ &= D(J_k, J_k, J_k)(K, I, I). \end{aligned}$$

Thus there is only one path component of signature $(1, k)$.

If n is even and $k = n/2$, J_k has only 2×2 blocks on its diagonal. Then $J_k K = \sigma J_k$, where $\sigma = \text{diag}(1, 1, \dots, 1, \sigma_2)$. Thus

$$\begin{aligned} D(J_k, J_k, J_k)(K, K, I) &= D(\sigma, \sigma, I)(J_k, J_k, J_k) \\ &= D(I, I, \sigma)(J_k, J_k, J_k) \\ &= D(J_k, J_k, J_k)(I, I, K), \end{aligned}$$

with similar formulas for the action of (K, I, K) and (I, K, K) on $D(J_k, J_k, J_k)$. One then sees that the $G(\mathbb{R})$ -orbit of $D(J_k, J_k, J_k)$ is the union of the $G^+(\mathbb{R})$ -orbits of $D(J_k, J_k, J_k)$, $D(J_k K, J_k, J_k)$, $D(J_k, J_k K, J_k)$, and $D(J_k, J_k, J_k K)$. To show that there are four path components of signature $(0, k)$, it remains to show that these four tensors lie in different components.

To this goal, consider a point P_0 of signature $(0, k)$, so that

$$P_0 = D(J_k, J_k, J_k)(g_1, g_2, g_3),$$

with $(g_1, g_2, g_3) \in G(\mathbb{R})$. Since $f_i(P_0; \mathbf{x}) \neq 0$, we also have $h_i(P_0; \mathbf{x}) \neq 0$. But by (4.5),

$$h_3(P_0; \mathbf{x}) = \det(g_1) \det(g_2) \det(J_k)^2 h_3(D(I, I, J_k); g_3 \mathbf{x}),$$

while a direct calculation shows that

$$h_3(D(I, I, J_k); \mathbf{x}) = \prod_{i=1}^k (x_{2i-1}^2 + x_{2i}^2).$$

Thus $h_3(P_0; \mathbf{x})$ is a nonzero polynomial in \mathbf{x} whose values are either nonnegative on all of \mathbb{R}^n or nonpositive, with similar statements valid for h_1, h_2 . But a straightforward continuity argument shows that along a path composed of points P with signature $(0, k)$ the polynomial $h_i(P; \mathbf{x})$, viewed as a function of \mathbf{x} , cannot pass between being nonnegative-valued and nonpositive-valued without being identically zero for some P . Since it is not identically zero at any point of signature $(0, k)$, on each path component it must be either nonnegative-valued for all P or nonpositive-valued for all P .

But since

$$\begin{aligned} h_3(D(J_k, J_k, J_k); \mathbf{x}) &= h_3(D(J_k, J_k, J_k)(I, I, K); \mathbf{x}) \\ &= -h_3(D(J_k, J_k, J_k)(K, I, I); \mathbf{x}) = -h_3(D(J_k, J_k, J_k)(I, K, I); \mathbf{x}) \end{aligned}$$

and

$$\begin{aligned} h_1(D(J_k, J_k, J_k); \mathbf{x}) &= h_1(D(J_k, J_k, J_k)(K, I, I); \mathbf{x}) \\ &= -h_1(D(J_k, J_k, J_k)(I, K, I); \mathbf{x}) = -h_1(D(J_k, J_k, J_k)(I, I, K); \mathbf{x}), \end{aligned}$$

we can conclude that the four points all lie in different path components, and so there are four path components of points of signature $(0, k)$. \square

Remark. Even in the well-studied case $n = 2$, the assertion of Theorem 6.3 seems to be a new result.

7. Application to a stochastic model. We consider here a discrete statistical model with a single hidden variable, in order to obtain a semialgebraic description of its set of probability distributions

As a graphical model, it is specified by a 3-leaf tree, as shown in Figure 7.1. The internal node of the tree, and each leaf, represent random variables, all with n states. The internal node variable is hidden (i.e., unobservable). The observed leaf variables are independent when conditioned on the state of the hidden one. The hidden variable thus provides an “explanation” of dependencies between the observed ones. This simple conditional independence model with a hidden variable is common in many statistical applications, and is variously called a hidden naive Bayes model, a latent class model, or a 3-leaf tree model, although often the four variables are allowed to have state spaces of different sizes.

In applications, the observed variables in this model might represent three characteristics (such as the results, + or −, of medical tests) measured on individuals in a population, while the hidden variable represents a “latent class” to which the individual belongs (such as whether the individual has or does not have a certain disease). The probabilities of the test outcomes depend on the disease condition, yet given an individual’s disease state, the results of the tests are independent of each other.

Given this model for fixed n , one can view a probability distribution arising from it as an $n \times n \times n$ tensor. A natural problem is to find a semialgebraic characterization of such tensors, that is, a collection of polynomial equalities and inequalities that precisely cut out the distributions arising from the model. The structure of the probability model ensures that the rank of the tensor is at most n and, moreover, that each summand in a decomposition as a sum of n rank-1 tensors has nonnegative entries. Understanding polynomial equalities holding on these distributions amounts to understanding the defining ideal of the variety V_n , work on which was reviewed in section 3. Inequalities holding on such tensors are much more poorly understood. While the existence of inequalities in the tensor entries that ensure the tensor arises from meaningful stochastic parameters (e.g., nonnegative) follows from general theory of real algebraic geometry, explicit inequalities have previously not been given for arbitrary n .

Our interest in the model is motivated by its appearance as the general Markov model in phylogenetics, where in the special case $n = 4$ it is used to model evolution of DNA sequences by base substitution. One might think of the unobserved variable as representing the base (A, C, T, or G) at a site in a sequence of an ancestral organism from which we have no data, and the observed variables as the state of the site in three currently extant descendants of it. Though trees with more than 3 leaves are, of course, essential for phylogenetic applications, in a related work [5] it is shown how

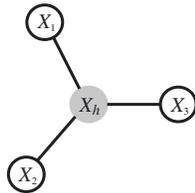


FIG. 7.1. A graphical model, in which the 3 leaves represent observed random variables X_1, X_2, X_3 , and the central node represents a hidden random variable X_h . As considered here, all variables are assumed to have n states. The structure of the graph indicates that the leaf variables are independent when conditioned on the hidden variable.

to extend a semialgebraic description for the 3-leaf tree to m -leaf trees.

Denoting this model by \mathcal{M}_n , we first describe its parametrization. With $[n] = \{1, 2, \dots, n\}$ as the state space of all random variables, let $\boldsymbol{\pi} = (\pi_1, \dots, \pi_n)$ denote the probability distribution vector for the hidden variable X_h , so that $\pi_i = \mathbb{P}(X_h = i)$. For the observed variables $X_i, i = 1, 2, 3$, an $n \times n$ stochastic matrix M_i has (j, k) -entry specifying the conditional probability $\mathbb{P}(X_i = k \mid X_h = j)$, so each row of each M_i sums to 1. The connection of this model to the tensor rank questions we study in this paper arises from the observation that a probability distribution for \mathcal{M}_n is specified by the $n \times n \times n$ tensor

$$P = \text{Diag}(\boldsymbol{\pi})(M_1, M_2, M_3).$$

The domain we consider for the parametrization map is specified by requiring that

1. $\boldsymbol{\pi}$ have strictly positive entries summing to 1,
2. the M_i have nonnegative entries and row sums of 1, and
3. the M_i are nonsingular.

For some statistical applications it is also natural to strengthen the second requirement so that the M_i have strictly positive entries; we thus comment on this situation as well in Proposition 7.1.

Note that a few trivial inequalities in the entries of P that must hold are obvious: Since $P \in \mathcal{M}_n$ is a probability distribution, its entries must be nonnegative. If one additionally assumes that the M_i have strictly positive entries, then P must have strictly positive entries as well.

For $n = 2$, a complete semialgebraic description of \mathcal{M}_2 has been given in two recent independent works [36, 19], using different approaches. In particular, in [36] the $2 \times 2 \times 2$ hyperdeterminant Δ plays a key role, though many statistically motivated ideas are also used. Here we give a semialgebraic description of \mathcal{M}_n for all $n \geq 2$, using the invariants developed in earlier sections that generalize Δ .

Recall a *principal minor* of a matrix is the determinant of a submatrix chosen with the same row and column indices. A *leading principal minor* is one for which these indices are $\{1, 2, 3, \dots, k\}$ for some k .

PROPOSITION 7.1. *An $n \times n \times n$ tensor P is in the image of the parametrization map for \mathcal{M}_n if and only if the following conditions hold:*

1. P is real, with nonnegative entries summing to 1.
2. For some (and hence all) i , P satisfies the commutation relations given by (3.1), and the polynomial $f_i(P; \mathbf{x})$ is not identically zero.
3. $\det(P *_i \mathbf{1}) \neq 0$ for all $i \in \{1, 2, 3\}$.
4. For at least one (and hence all) of the following matrices, all leading principal minors are strictly positive:

$$(7.1) \quad \begin{aligned} & \det(P *_1 \mathbf{1})(P *_2 \mathbf{1}) \text{adj}(P *_1 \mathbf{1})(P *_3 \mathbf{1})^T, \\ & \det(P *_2 \mathbf{1})(P *_1 \mathbf{1}) \text{adj}(P *_2 \mathbf{1})(P *_3 \mathbf{1}), \\ & \det(P *_3 \mathbf{1})(P *_1 \mathbf{1})^T \text{adj}(P *_3 \mathbf{1})(P *_2 \mathbf{1}). \end{aligned}$$

5. For all $1 \leq l \leq n$, all principal minors of the three matrices

$$(7.2) \quad \begin{aligned} & \det(P *_1 \mathbf{1})(P *_2 \mathbf{1}) \text{adj}(P *_1 \mathbf{1})(P *_3 \mathbf{e}_l)^T, \\ & \det(P *_1 \mathbf{1})(P *_2 \mathbf{e}_l) \text{adj}(P *_1 \mathbf{1})(P *_3 \mathbf{1})^T, \\ & \det(P *_2 \mathbf{1})(P *_1 \mathbf{e}_l) \text{adj}(P *_2 \mathbf{1})(P *_3 \mathbf{1}) \end{aligned}$$

are nonnegative.

Here $\text{adj}(M)$ denotes the classical adjoint of a matrix M .

If parameters of the model are restricted so that entries of M_i are strictly positive, then in condition 5 one should replace “principal minors” by “leading principal minors” and “nonnegative” by “positive.”

Note that the only equality constraints in the theorem are those in conditions 1 and 2. In particular, a full set of generators of the ideal $I(V_n)$ is not used (when $n \geq 3$) in this semialgebraic description of the model.

Our proof will repeatedly use the following well-known classical result on matrices defining quadratic forms.

THEOREM 7.2 (Sylvester’s theorem). *Let A be an $n \times n$ real symmetric matrix and $Q(\mathbf{v}) = \mathbf{v}^T A \mathbf{v}$ the associated quadratic form on \mathbb{R}^n . Then the following hold:*

1. *Q is positive definite if and only if all leading principal minors of A are strictly positive.*
2. *Q is positive semidefinite if and only if all principal minors of A are nonnegative.*

Proof of Proposition 7.1. We first discuss the necessity of these conditions. The necessity of condition 1 is clear. Condition 2 holds by Theorem 4.1, since

$$P = \text{Diag}(\boldsymbol{\pi})(M_1, M_2, M_3) = D(\text{diag}(\boldsymbol{\pi})M_1, M_2, M_3)$$

shows that P is in the $G(\mathbb{C})$ -orbit of D .

For condition 3, observe that $P *_i \mathbf{1}$ is the marginalization of the distribution to two observed variables; thus one sees that

$$P *_i \mathbf{1} = M_j^T \text{diag}(\boldsymbol{\pi}) M_k$$

for distinct i, j, k . Since $\boldsymbol{\pi}$ has positive entries, and M_j, M_k are nonsingular, the determinant of this matrix is nonzero.

Condition 4 can be restated, after dividing the first formula of (7.1) by the positive number $\det(P *_1 \mathbf{1})^2$, as asserting the positivity of leading principal minors of

$$(P *_2 \mathbf{1})(P *_1 \mathbf{1})^{-1}(P *_3 \mathbf{1})^T$$

and two similar expressions. But expressed in terms of parameters, this is

$$(M_1^T \text{diag}(\boldsymbol{\pi}) M_3)(M_2^T \text{diag}(\boldsymbol{\pi}) M_3)^{-1}(M_1^T \text{diag}(\boldsymbol{\pi}) M_2)^T = M_1^T \text{diag}(\boldsymbol{\pi}) M_1.$$

This symmetric matrix, and similar matrices obtained from the other expressions, define positive definite quadratic forms because $\boldsymbol{\pi}$ has positive entries. Thus by Sylvester’s theorem, all their principal minors are positive.

A similar argument shows the necessity of condition 5. For instance, letting \mathbf{r}_l^3 be the vector whose entries are the products $\pi_i M_3(i, l)$, one sees that

$$\det(P *_1 \mathbf{1})(P *_2 \mathbf{1}) \text{adj}(P *_1 \mathbf{1})(P *_3 \mathbf{e}_l)^T = \det(P *_1 \mathbf{1})^2 M_1^T \text{diag}(\mathbf{r}_l^3) M_1.$$

Since the entries of M_3 are nonnegative, this matrix defines a positive semidefinite quadratic form, and thus by Sylvester’s theorem has nonnegative principal minors. If the entries of the M_i are positive, then this matrix defines a positive definite form and thus has positive leading principal minors.

Turning to sufficiency, assume that conditions 1–5 are met by a tensor P . By Theorem 4.1, condition 2 implies $P = D(g_1, g_2, g_3)$ for some $g_i \in GL(n, \mathbb{C})$. Moreover, by the realness of P in condition 1, from Lemma 6.1 we also know that any

complex entries in the g_i occur in complex conjugate rows. Our goal is to modify this expression, so the g_i are replaced by stochastic matrices, and D by a diagonal tensor with positive entries.

Letting $\mathbf{s}_i = g_i \mathbf{1}$ be the vector of row sums of g_i , we have

$$\begin{aligned} P *_1 \mathbf{1} &= g_2^T \operatorname{diag}(\mathbf{s}_1) g_3, \\ P *_2 \mathbf{1} &= g_1^T \operatorname{diag}(\mathbf{s}_2) g_3, \\ P *_3 \mathbf{1} &= g_1^T \operatorname{diag}(\mathbf{s}_3) g_2. \end{aligned}$$

Thus the nonvanishing of the determinants of these matrices by condition 3 tells us that the row sums are all nonzero. Letting $M_i = \operatorname{diag}(\mathbf{s}_i)^{-1} g_i$ and $\boldsymbol{\pi}$ be the vector of entrywise products of $\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3$, we thus have $P = \operatorname{Diag}(\boldsymbol{\pi})(M_1, M_2, M_3)$. Here each M_i has unit row sums, $\boldsymbol{\pi}$ has nonzero entries, and

$$\begin{aligned} P *_1 \mathbf{1} &= M_2^T \operatorname{diag}(\boldsymbol{\pi}) M_3, \\ P *_2 \mathbf{1} &= M_1^T \operatorname{diag}(\boldsymbol{\pi}) M_3, \\ P *_3 \mathbf{1} &= M_1^T \operatorname{diag}(\boldsymbol{\pi}) M_2. \end{aligned}$$

Since P is real, these expressions are as well, although we have not yet shown that M_i and $\boldsymbol{\pi}$ have real entries. Nonetheless, all M_i have the same number of conjugate (nonreal) pairs of rows, in corresponding positions, with the corresponding entries of $\boldsymbol{\pi}$ also conjugate (though possibly real).

Now substituting the above expressions for marginalizations in the three expressions in condition 4, we have that they simplify to

$$\begin{aligned} \det(P *_1 \mathbf{1})^2 M_1^T \operatorname{diag}(\boldsymbol{\pi}) M_1, \\ \det(P *_2 \mathbf{1})^2 M_2^T \operatorname{diag}(\boldsymbol{\pi}) M_2, \\ \det(P *_3 \mathbf{1})^2 M_3^T \operatorname{diag}(\boldsymbol{\pi}) M_3. \end{aligned}$$

This shows that $M_i^T \operatorname{diag}(\boldsymbol{\pi}) M_i$ is real for each i . We now argue that if M_i is not real, then the quadratic form Q_i associated to $M_i^T \operatorname{diag}(\boldsymbol{\pi}) M_i$ is not positive definite. To that end, suppose that two rows (say, the first two) of M_i are complex conjugates and thus, by Lemma 6.1, of the form

$$\mathbf{m}_i^1 = \mathbf{r}_1 + i\mathbf{r}_2, \quad \mathbf{m}_i^2 = \overline{\mathbf{m}_i^1} = \mathbf{r}_1 - i\mathbf{r}_2, \quad \mathbf{r}_i \in \mathbb{R}^n \setminus \{\mathbf{0}\},$$

and that the corresponding entries of $\boldsymbol{\pi}$ are $\pi_1, \pi_2 = \overline{\pi_1}$. Then for any real vector \mathbf{v} orthogonal to the real and imaginary parts of the other rows of M_i , evaluating the quadratic form at \mathbf{v} yields

$$Q_i(\mathbf{v}) = \pi_1(\mathbf{m}_i^1 \cdot \mathbf{v})^2 + \overline{\pi_1}(\overline{\mathbf{m}_i^1} \cdot \mathbf{v})^2.$$

If we additionally choose \mathbf{v} to be orthogonal to \mathbf{r}_2 , but not to \mathbf{r}_1 , then

$$Q_i(\mathbf{v}) = \pi_1(\mathbf{r}_1 \cdot \mathbf{v})^2 + \overline{\pi_1}(\mathbf{r}_1 \cdot \mathbf{v})^2 = 2\Re(\pi_1)(\mathbf{r}_1 \cdot \mathbf{v})^2.$$

Positive definiteness of Q_i would thus imply $\Re(\pi_1) > 0$. However, if we instead choose \mathbf{v} to be orthogonal to \mathbf{r}_1 , but not to \mathbf{r}_2 , then

$$Q_i(\mathbf{v}) = \pi_1(i\mathbf{r}_2 \cdot \mathbf{v})^2 + \overline{\pi_1}(i\mathbf{r}_2 \cdot \mathbf{v})^2 = -2\Re(\pi_1)(\mathbf{r}_2 \cdot \mathbf{v})^2,$$

so positive definiteness would imply $\Re(\pi_1) < 0$. Thus if M_i were not real, then Q_i would not be positive definite.

But if Q_i is not positive definite, by Sylvester’s theorem, the positivity of leading principal minors asserted in condition 4 must be violated. Thus condition 4 implies that at least one of the M_i is real, so all are by Lemma 6.1. Applying Sylvester’s theorem again to the positive definite form Q_i then implies that π has real positive entries.

Finally, condition 5 ensures that the entries of the M_i are nonnegative. For instance,

$$(7.3) \quad \det(P *_1 \mathbf{1})(P *_2 \mathbf{1}) \operatorname{adj}(P *_1 \mathbf{1})(P *_3 \mathbf{e}_j)^T = \det(P *_1 \mathbf{1})^2 M_1^T \operatorname{diag}(\pi) \operatorname{diag}(\tilde{\mathbf{m}}_3^j) M_1,$$

where \tilde{m}_3^j is the j th column of M_3 . Thus all principal minors of this matrix being nonnegative implies that the associated quadratic form is positive semidefinite and thus that \tilde{m}_3^j has nonnegative entries. To instead ensure that these entries are strictly positive, we require that the quadratic form be positive definite, and thus that all leading principal minors be positive. \square

For further work in this direction, we direct the reader to [5].

8. Application to “rank jumping.” It is well known that the limit of a sequence of tensors of a fixed rank $r > 1$ may be strictly larger than r (that is, tensor rank, unlike matrix rank, is not upper semicontinuous). This “rank jumping” is responsible for the fact that a given tensor may not have a best approximation by a tensor of fixed lower rank and can thus be of concern in applied settings.

A tensor that is the limit of tensors of rank r , but not of smaller rank, is said to have *border rank* r . Thus the border rank of a tensor is always less than or equal to its rank.

For instance, while the tensors of complex rank 2 are dense among the $2 \times 2 \times 2$ tensors, there is a unique $G(\mathbb{C})$ -orbit of rank-3 tensors [10], which therefore have border rank 2. An orbit representative, called the Werner tensor in the physics literature, is usually taken as

$$\mathbf{e}_2 \otimes \mathbf{e}_1 \otimes \mathbf{e}_1 + \mathbf{e}_1 \otimes \mathbf{e}_2 \otimes \mathbf{e}_1 + \mathbf{e}_1 \otimes \mathbf{e}_1 \otimes \mathbf{e}_2.$$

For our purposes, it is more convenient to apply a permutation in the second index, so that its 3-slices become

$$W = \left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \right]$$

and thus have the form described in Proposition 3.3. One may express W as an explicit limit of rank-2 tensors using a difference quotient [10]. This difference quotient construction generalizes to other formats, to produce simple examples of tensors whose rank is larger than their border rank.

Proposition 3.3 suggests a different way of obtaining W and many other tensors whose rank exceeds their border rank. Our goal in this section is to provide some explicit examples.

In the $n = 3$ case, consider the tensor given by 3-slices as

$$K_3 = \left[\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \right]$$

and its perturbation as

$$K_{3,\epsilon} = \left[\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 \\ 0 & \epsilon & 1 \\ 0 & 0 & 2\epsilon \end{pmatrix}, \begin{pmatrix} 0 & \epsilon & 1 \\ 0 & \epsilon^2 & 3\epsilon \\ 0 & 0 & 4\epsilon^2 \end{pmatrix} \right].$$

Both tensors are 3-slice nonsingular and have commuting slices, since for each the third slice is the square of the second.

For arbitrary n , one can similarly construct K_n with slices whose entries are all zeros except for successive superdiagonals of 1's. Perturbing the diagonal of the second slice by adding $(0, \epsilon, 2\epsilon, 3\epsilon, \dots, (n - 1)\epsilon)$, we can perturb the other slices to be appropriate powers of the perturbed second slice, so that we obtain $K_{n,\epsilon}$ with all 3-slices commuting. The matrix $Z_{n,\epsilon}$ of diagonals of the slices of $K_{n,\epsilon}$ is then a Vandermonde matrix and hence nonsingular for $\epsilon \neq 0$.

Now $K_{n,\epsilon}$ meets the hypotheses of Proposition 3.3 and has slices already upper-triangularized. Moreover, since $Z_{n,\epsilon}$ is nonsingular for $\epsilon \neq 0$, it follows that $K_{n,\epsilon} \in \mathcal{D}(\mathbb{C})$.

Since K_n is in the closure of all $K_{n,\epsilon}$, we see that $K \in V_n$ has border rank at most n . Since the multilinear rank of K_n is (n, n, n) , it cannot have border rank less than n , so its border rank is exactly n . Since $f_3(K_n; \mathbf{x}) = 0$, Theorem 4.1 shows that $K_n \notin \mathcal{D}(\mathbb{C})$. Since this orbit is precisely the tensors of rank n and multilinear rank (n, n, n) , this implies that the tensor rank of K_n must be strictly greater than n .

We now determine the rank precisely.

PROPOSITION 8.1. *For any $n > 0$, K_n has border rank n and rank $2n - 1$, over \mathbb{C} .*

Proof. The fact that K_n has border rank n has been discussed.

To show that the rank is at most $2n - 1$ we give an explicit representation, suggested by A. Jensen. We work with a more symmetric tensor K'_n , obtained by acting on K_n by a permutation in the second index, reversing the order of the columns of each slice. For example,

$$K'_3 = \left[\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \right].$$

In general, K'_n will be the $n \times n \times n$ tensor of all zeros, except for 1's in the (i, j, k) position when $i + j + k = n + 2$.

Let ζ denote a primitive $(2n - 1)$ th root of unity. Let $\mathbf{v}_l = (\zeta^l, \zeta^{2l}, \dots, \zeta^{nl})$. Then we claim that

$$K'_n = \sum_{l=1}^{2n-1} \frac{1}{2n-1} \zeta^{l(n-3)} \mathbf{v}_l \otimes \mathbf{v}_l \otimes \mathbf{v}_l,$$

and thus K'_n has rank at most $2n - 1$. Indeed, the (i, j, k) entry of this sum is

$$\frac{1}{2n-1} \sum_{l=1}^{2n-1} \zeta^{l(i+j+k+n-3)} = \begin{cases} 1 & \text{if } (2n-1) \mid (i+j+k+n-3), \\ 0 & \text{otherwise.} \end{cases}$$

Since $n \leq (i + j + k + n - 3) \leq 4n - 3$, the nonzero entries occur only when $i + j + k + n - 3 = 2n - 1$, i.e., when $i + j + k = n + 2$.

To see that the rank is at least $2n - 1$, suppose K_n could be expressed as

$$K_n = \sum_{l=1}^{2n-2} \mathbf{u}_l \otimes \mathbf{v}_l \otimes \mathbf{w}_l,$$

with $\mathbf{u}_i, \mathbf{v}_i, \mathbf{w}_i \in \mathbb{C}^n$. Since the $12|3$ flattening of K_n has rank n , the \mathbf{w}_i must span \mathbb{C}^n . Thus without loss of generality we may assume that $\mathcal{B} = \{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_n\}$ is a basis for \mathbb{C}^n . Let $\mathcal{B}^* = \{\mathbf{w}_1^*, \mathbf{w}_2^*, \dots, \mathbf{w}_n^*\}$ be the dual basis. Then for any $i \in \{1, 2, \dots, n\}$, \mathbf{w}_i^* annihilates at least $n - 1$ of the \mathbf{w}_j , so $K_n *_3 \mathbf{w}_i^*$ is a matrix of rank at most $(2n - 2) - (n - 1) = n - 1$. But one sees from the explicit form of K_n that $K_n *_3 \mathbf{w}_i^*$ can have rank at most $n - 1$ only if \mathbf{w}_i^* has first coordinate 0. This contradicts that \mathcal{B}^* is a basis. \square

One can construct many other examples of rank jumping by considering variants of the arguments above using different Jordan block structures of the slices. For example,

$$L = \left[\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right]$$

can be perturbed to

$$L_\epsilon = \left[\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 \\ 0 & \epsilon & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right] \in \mathcal{D}(\mathbb{C}).$$

Here one can see that L has tensor rank at most 4 (by subtracting the third slice from the first). Since it also has multilinear rank $(3, 3, 3)$ and by Theorem 4.1 is not in $\mathcal{D}(\mathbb{C})$, its tensor rank must be exactly 4.

Finally, we note that the maximal \mathbb{C} -rank of an $n \times n \times n$ tensor of border rank n for $n = 2$ is well known to be 3. For $n = 3$, it is claimed in [21] that the maximal rank is 5. The tensor K_n achieves these bounds in both cases. We know of no examples of $n \times n \times n$ tensors of border rank n whose rank exceeds $2n - 1$, the rank of K_n . It has been conjectured by one of the authors (JAR; see [6]) that no such tensors exist.

It should be noted that the tensors K_n described in this section are similar to those given in Theorem 5.6 of [1] of size $n \times n \times (\lfloor \log_2 n \rfloor + 1)$, whose rank is $2n - 1$ when $n = 2^k$ and whose border rank has been shown to be n [22]. (When $n \neq 2^k$, the rank of these tensors is slightly smaller than $2n - 1$.) Related examples are also given in Corollary 5.7 of [1] of tensors of size $n \times (n + 1) \times (n + 1)$ and rank approximately $3n$. However, the border rank of these tensors has only been shown to be bounded above by approximately $2n$ when $n = 2^k$ [22], with the precise border rank unknown. Thus it is unclear what the gap between rank and border rank is for these last examples.

Acknowledgments. Much of the research in this work was conducted while ESA and JAR were in residence at the University of Tasmania as visiting scholars; they thank the university and the School of Mathematics and Physics for their hospitality. ESA and JAR also thank the Institut Mittag-Leffler and the organizers of and participants in its Spring 2011 Program in Algebraic Geometry with a View toward Applications for stimulating conversations. They thank Allesandra Bernardi especially for many helpful clarifications in discussions, and Anders Jensen for his contribution to the argument for Proposition 8.1.

REFERENCES

- [1] B. ALEXEEV, M. FORBES, AND J. TSIMERMAN, *Tensor rank: Some lower and upper bounds*, in Proceedings of the IEEE 26th Annual Conference on Computational Complexity, 2011, pp. 283–291.
- [2] E. S. ALLMAN, P. D. JARVIS, J. A. RHODES, AND J. G. SUMNER, Supplementary materials, 2013, <http://www.maths.utas.edu.au/People/Holland/software.htm>.
- [3] E. S. ALLMAN AND J. A. RHODES, *Phylogenetic invariants for the general Markov model of sequence mutation*, Math. Biosci., 186 (2003), pp. 113–144.
- [4] E. S. ALLMAN AND J. A. RHODES, *Phylogenetic ideals and varieties for the general Markov model*, Adv. in Appl. Math., 40 (2008), pp. 127–148.
- [5] E. S. ALLMAN, J. A. RHODES, AND A. TAYLOR, *A Semialgebraic Description of the General Markov Model on Phylogenetic Trees*, preprint, arXiv:1212.1200v1 [q-bio.PE], 2012.
- [6] E. BALLICO AND A. BERNARDI, *Stratification of the Fourth Secant Variety of Veronese Variety via Symmetric Rank*, preprint, arXiv:1005.3465v3 [math.AG], 2011.
- [7] D. J. BATES AND L. OEDING, *Toward a salmon conjecture*, Exp. Math., 20 (2011), pp. 358–370.
- [8] A. CAYLEY, *On the theory of linear transformations*, Cambridge Math. J., 4 (1845), pp. 193–209.
- [9] V. COFFMAN, J. KUNDU, AND W. K. WOOTTERS, *Distributed entanglement*, Phys. Rev. A, 61 (2000), 052306.
- [10] V. DE SILVA AND L.-H. LIM, *Tensor rank and the ill-posedness of the best low-rank approximation problem*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 1084–1127.
- [11] H. DERKSEN AND G. KEMPER, *Computational Invariant Theory*, Invariant Theory Algebr. Transform. Groups, I, Encyclopaedia Math. Sci. 130, Springer-Verlag, Berlin, 2002.
- [12] S. FRIEDLAND, *On tensors of border rank l in $\mathbb{C}^{m \times n \times l}$* , Linear Algebra Appl., 438 (2013), pp. 713–737.
- [13] S. FRIEDLAND AND E. GROSS, *A proof of the set-theoretic version of the salmon conjecture*, J. Algebra, 356 (2012), pp. 374–379.
- [14] W. FULTON AND J. HARRIS, *Representation Theory. A First Course*, Grad. Texts in Math. 129, Springer-Verlag, New York, 1991.
- [15] L. D. GARCIA, M. STILLMAN, AND B. STURMFELS, *Algebraic geometry of Bayesian networks*, J. Symbolic Comput., 39 (2005), pp. 331–355.
- [16] I. M. GELFAND, M. M. KAPRANOV, AND A. V. ZELEVINSKY, *Discriminants, Resultants and Multidimensional Determinants*, Mod. Birkhäuser Class., Birkhäuser Boston, Boston, MA, 2008.
- [17] R. GOODMAN AND N. R. WALLACH, *Representations and Invariants of the Classical Groups*, Cambridge University Press, Cambridge, UK, 2003.
- [18] J. H. GRACE AND A. YOUNG, *The algebra of invariants*, Cambridge Libr. Collect., Cambridge University Press, Cambridge, UK, 2010.
- [19] S. KLAERE AND V. LIEBSCHER, *An algebraic analysis of the two state Markov model on tripod trees*, Math. Biosci., 237 (2012), pp. 38–48.
- [20] J. B. KRUSKAL, *Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics*, Linear Algebra Appl., 18 (1977), pp. 95–138.
- [21] J. B. KRUSKAL, *Rank, decomposition, and uniqueness for 3-way and N -way arrays*, in Multiway Data Analysis (Rome, 1988), North-Holland, Amsterdam, 1989, pp. 7–18.
- [22] J. LANDSBERG, *Explicit Tensors in $C^n \otimes C^n \otimes C^n$ of Border Rank at Least $2n - 1$* , preprint, arXiv:1209.1664v1 [cs.CC], 2012.
- [23] J. M. LANDSBERG, *Geometry and the complexity of matrix multiplication*, Bull. Amer. Math. Soc. (N.S.), 45 (2008), pp. 247–284.
- [24] D. E. LITTLEWOOD, *The Theory of Group Characters and Matrix Representations of Groups*, AMS Chelsea Publishing, Providence, RI, 2006.
- [25] I. G. MACDONALD, *Symmetric Functions and Hall Polynomials*, The Clarendon Press, Oxford University Press, New York, 1979.
- [26] K. O. NG, *The classification of $(3, 3, 3)$ trilinear forms*, J. Reine Angew. Math., 468 (1995), pp. 49–75.
- [27] P. J. OLVER, *Classical Invariant Theory*, London Math. Soc. Stud. Texts 44, Cambridge University Press, Cambridge, UK, 1999.
- [28] J. PEARL AND M. TARSI, *Structuring causal trees*, J. Complexity, 2 (1986), pp. 60–77.
- [29] J. A. RHODES, *A concise proof of Kruskal’s theorem on tensor decomposition*, Linear Algebra Appl., 432 (2010), pp. 1818–1824.
- [30] A. SEIDENBERG, *A new decision method for elementary algebra*, Ann. of Math. (2), 60 (1954), pp. 365–374.

- [31] V. STRASSEN, *Rank and optimal computation of generic tensors*, Linear Algebra Appl., 52/53 (1983), pp. 645–685.
- [32] J. G. SUMNER, M. A. CHARLESTON, L. S. JERMIN, AND P. D. JARVIS, *Markov invariants, plethysms, and phylogenetics*, J. Theoret. Biol., 253 (2008), pp. 601–615.
- [33] J. G. SUMNER AND P. D. JARVIS, *Using the tangle: A consistent construction of phylogenetic distance matrices*, Math. Biosci., 204 (2006), pp. 49–67.
- [34] A. TARSKI, *A Decision Method for Elementary Algebra and Geometry*, 2nd ed., University of California Press, Berkeley, Los Angeles, CA, 1951.
- [35] B. G. WYBOURNE, *Schur: An interactive programme for calculating properties of Lie groups*, version 6.03, <http://sourceforge.net/projects/schur> (2004).
- [36] P. ZWIERNIK AND J. Q. SMITH, *Implicit inequality constraints in a binary tree model*, Electron. J. Stat., 5 (2011), pp. 1276–1312.