CrossMark

CAMBRIDGE
UNIVERSITY PRESS

RESEARCH ARTICLE

Maximum likelihood estimation for tensor normal models via castling transforms

Harm Derksen [ID][1], Visu Makam [ID][2] and Michael Walter [ID][3,4]

[1]Department of Mathematics, Northeastern University, 567 Lake Hall, 360 Huntington Ave, Boston, MA 02115, USA;
E-mail: ha.derksen@northeastern.edu.
[2]Radix Trading Europe B. V., Strawinskylaan 1217, Amsterdam, 1082 MK, Netherlands; E-mail: visu@umich.edu.
[3]Faculty of Computer Science, Ruhr-Universität Bochum, Universitätsstr. 150, 44801 Bochum, Germany;
E-mail: michael.walter@rub.de.
[4]Korteweg-de Vries Institute for Mathematics, Institute for Theoretical Physics, Institute for Logic, Language and Computation,
QuSoft, University of Amsterdam, Science Park 105-107, 1098 XG Amsterdam, The Netherlands.

Abstract

In this paper, we study sample size thresholds for maximum likelihood estimation for tensor normal models. Given the model parameters and the number of samples, we determine whether, almost surely, (1) the likelihood function is bounded from above, (2) maximum likelihood estimates (MLEs) exist, and (3) MLEs exist uniquely. We obtain a complete answer for both real and complex models. One consequence of our results is that almost sure boundedness of the log-likelihood function guarantees almost sure existence of an MLE. Our techniques are based on invariant theory and castling transforms.

Contents

1. Introduction

A family of probability distributions is called a *statistical model*. Maximum likelihood estimation is a method of estimating the true probability distribution as the one that maximizes the likelihood of the observed data. The probability distribution (or often the point in an associated parameter space) that maximizes the likelihood is called a *maximum likelihood estimate (MLE)*. One important problem is to understand the minimal number of samples required such that, almost surely, (1) the likelihood function is bounded from above, (2) MLEs exist, and (3) there is a unique MLE. Surprising connections

between sample size thresholds for a class of models called Gaussian group models and stability notions in invariant theory were recently discovered in [2]. In this paper, we study sample size thresholds for *tensor normal models*, which fall under the purview of Gaussian group models and are hence amenable to techniques from invariant theory. The setting of invariant theory that relates to tensor normal models are the so-called tensor actions: that is, the natural action of the group $\mathrm{SL}_{d_1} \times \mathrm{SL}_{d_2} \times \cdots \times \mathrm{SL}_{d_k}$ on $\mathbb{F}^{d_1} \otimes \mathbb{F}^{d_2} \otimes \cdots \otimes \mathbb{F}^{d_k}$, where $\mathbb{F}$ is the underlying field (either $\mathbb{R}$ or $\mathbb{C}$) and SL_{d_i} denotes the group of $d_i \times d_i$ matrices with determinant one.

Tensor normal models are statistical models consisting of multivariate Gaussian distributions whose concentration matrix is a Kronecker (or tensor) product of several matrices. These are particularly useful in studying data that naturally occurs as multidimensional arrays. Examples include wood density in given growth rings and directions at several heights in a tree trunk [24], monitoring of a vector of physiological variables in different organs over multiple days [34] and 3-dimensional spatial glucose content data [27]. Tensors are also ubiquitous in big data applications.

A special case of tensor normal models is the *matrix normal model*, where the concentration matrix is a Kronecker product of exactly two matrices. Sample size thresholds for matrix normal models have been investigated in [14, 32, 37, 13, 36, 2, 12]. In particular, a complete answer for matrix normal models was obtained in [12] with techniques from quiver representations. We do not use quiver representations in this paper; instead, we use castling transforms and results on stabilisers in general position. It is worth mentioning that the invariant theory for tensor actions with two tensor factors (which corresponds to the matrix normal models) is well understood, and we have efficient algorithms (see [18, 8, 19, 20, 9, 10, 1]), whereas the invariant theory becomes significantly more difficult for three and more tensor factors (see [6, 7, 11] for more details).

To find the MLE, one can use the so-called flip-flop algorithm [14, 25, 26, 42] for matrix normal models and its natural generalisations to tensor normal models; it is closely related to a recent alternating minimisation algorithm in the invariant theory of tensor actions [2, 6, 16]. In general, MLEs for Gaussian group models can be found using the geodesic optimisation algorithms in [7].

A separate motivation for studying the questions in this paper comes from quantum information. Here tensors describe the states of a quantum mechanical system, and our invariant theoretic results characterise the existence of states with certain prescribed marginals; see [22, 40, 41, 4, 5] for details.

1.1. Tensor normal models

Let $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. Let PD_n denote the cone of positive definite $n \times n$ matrices with entries in $\mathbb{F}$. For an n-dimensional centred Gaussian distribution (circularly symmetric when $\mathbb{F} = \mathbb{C}$) with concentration matrix $\Psi \in \mathrm{PD}_n$, the density function is defined as

$$f_\Psi(y) = \begin{cases} \det\left(\frac{\Psi}{2\pi}\right)^{\frac{1}{2}} \cdot e^{-\frac{1}{2} y^\dagger \Psi y}, & \text{if } \mathbb{F} = \mathbb{R}; \\ \det\left(\frac{\Psi}{\pi}\right) \cdot e^{-y^\dagger \Psi y}, & \text{if } \mathbb{F} = \mathbb{C}. \end{cases}$$

Note that $y^\dagger$ denotes the adjoint (conjugate transpose) of y.

Given a subset $\mathcal{M} \subseteq \mathrm{PD}_n$, we define the corresponding *Gaussian model* as the statistical model consisting of the distributions with concentration matrix $\Psi \in \mathcal{M}$. Then the likelihood function $L_Y : \mathcal{M} \to \mathbb{R}$ is, for m samples specified by an m-tuple $Y = (Y_1, \ldots, Y_m) \in (\mathbb{F}^n)^m$, given by

$$L_Y(\Psi) = \prod_{i=1}^{m} f_\Psi(Y_i) = \begin{cases} \det\left(\frac{\Psi}{2\pi}\right)^{m/2} \cdot e^{-\frac{1}{2} \sum_{i=1}^{m} Y_i^\dagger \Psi Y_i}, & \text{if } \mathbb{F} = \mathbb{R}; \\ \det\left(\frac{\Psi}{\pi}\right)^{m} \cdot e^{-\sum_{i=1}^{m} Y_i^\dagger \Psi Y_i}, & \text{if } \mathbb{F} = \mathbb{C}. \end{cases}$$

For both $\mathbb{F} = \mathbb{R}$ and $\mathbb{F} = \mathbb{C}$, the log-likelihood function can be written as (up to an additive constant and multiplicative constant)

$$l_Y(\Psi) = \frac{m}{2}\log\det(\Psi) - \frac{1}{2}\mathrm{Tr}\left(\Psi\sum_{i=1}^{m}Y_iY_i^{\dagger}\right). \tag{1.1}$$

A *maximum likelihood estimate (MLE)* given Y is a concentration matrix $\hat{\Psi} \in \mathcal{M}$ that maximises the likelihood of observing the data Y: that is, $l_Y(\hat{\Psi}) \geq l_Y(\Psi)$ for all $\Psi \in \mathcal{M}$. For an MLE to exist, it is therefore necessary (but not necessarily sufficient) that l_Y is bounded from above. Even when they exist, MLEs need not be unique.

For $d_1, \ldots, d_k \in \mathbb{Z}_{>0}$, the Gaussian model $\mathcal{M}(d_1, \ldots, d_k) = \{\Psi_1 \otimes \Psi_2 \otimes \cdots \otimes \Psi_k \mid \Psi_i \in \mathrm{PD}_{d_i}\} \subseteq \mathrm{PD}_n$ (where $n = d_1 d_2 \cdots d_k$) is called a *tensor normal model*. When we want to differentiate between the real and the complex model, we will write $\mathcal{M}_{\mathbb{R}}(d_1, \ldots, d_k)$ and $\mathcal{M}_{\mathbb{C}}(d_1, \ldots, d_k)$, respectively. For the tensor normal model $\mathcal{M}(d_1, \ldots, d_k)$, a sample can be thought of not only as a vector of size n but also as a k-tensor with local dimensions $d_1, d_2, \ldots, d_k$. The latter viewpoint will be particularly useful. Accordingly, we define $\mathbb{F}^{d_1, \ldots, d_k} := \mathbb{F}^{d_1} \otimes \mathbb{F}^{d_2} \otimes \cdots \otimes \mathbb{F}^{d_k}$. Then a sample for the tensor normal model $\mathcal{M}(d_1, \ldots, d_k)$ is simply a point in the tensor space $\mathbb{F}^{d_1, \ldots, d_k}$. We also write $\mathbb{F}^{d_1, \ldots, d_k; m}$ for $(\mathbb{F}^{d_1, \ldots, d_k})^{\oplus m}$.

1.2. Main results on sample size thresholds

Generalising the quantity $R(d_1, \ldots, d_k)$ defined in [4], we consider

$$R(d_1, \ldots, d_k; m) := m\prod_{i=1}^{k}d_i + \sum_{n=1}^{k}(-1)^n\sum_{1\leq i_1<\ldots<i_n\leq k}\gcd(d_{i_1}, \ldots, d_{i_n})^2,$$

as well as the following two quantities:

$$\Delta(d_1, \ldots, d_k; m) := m\prod_{i=1}^{k}d_i - 1 - \sum_{i=1}^{k}(d_i^2 - 1), \qquad g_{\max}(d_1, \ldots, d_k) := \max_{i<j}\gcd(d_i, d_j).$$

By convention, $g_{\max}(d) = 1$ for any $d \in \mathbb{Z}_{>0}$. Then all three quantities are invariant under leaving out dimensions equal to one. The following theorem shows that these quantities precisely predict the almost sure behavior of the MLE. By almost sure, we mean the stated property holds for all Y away from a subset of $\mathbb{F}^{d_1, \ldots, d_k; m} \cong (\mathbb{F}^n)^m$ of Lebesgue measure zero.

Theorem 1.1. *Let $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. Consider m samples $Y = (Y_1, \ldots, Y_m)$ of the tensor normal model $\mathcal{M}(d_1, \ldots, d_k)$. Let $R = R(d_1, \ldots, d_k; m)$, $\Delta = \Delta(d_1, \ldots, d_k; m)$ and $g_{\max} = g_{\max}(d_1, \ldots, d_k)$. Then*

1. *If $R > 0$, then almost surely an MLE exists. Furthermore:*
   - *If $m \geq 2$, the MLE is almost surely unique if and only if $R > g_{\max}^2$ or $g_{\max} = 1$.*
   - *If $m = 1$, the MLE is almost surely unique if and only if $\Delta \geq -1$.*
2. *If $R = 0$, then almost surely an MLE exists. It is almost surely unique if and only if $g_{\max} = 1$.*
3. *If $R < 0$, then the likelihood function is always unbounded from above.*

Remark 1.2. It was conjectured in [13] and proved in [12] that for matrix normal models (tensor normal models with $k = 2$), almost sure boundedness of the log-likelihood function implies almost sure existence of an MLE. Theorem 1.1 implies that the same holds for all tensor normal models.

From Theorem 1.1, we can extract the following result. Let us denote by mlt_b (respectively mlt_e, mlt_u) the smallest integer m_0 such that, for all $m \geq m_0$, the log-likelihood function for $Y \in \mathbb{F}^{d_1, \ldots, d_k; m}$ is almost surely bounded from above (respectively MLEs exist, the MLE exists uniquely).

Corollary 1.3. *Let* $\mathbb{F} = \mathbb{R}$ *or* $\mathbb{C}$. *Consider the tensor normal model* $\mathcal{M}(d_1, \ldots, d_k)$. *Without loss of generality, assume* $2 \leq d_1 \leq d_2 \leq \cdots \leq d_k$, *and assume* $k \geq 3$. *Let* $r = \frac{d_k}{d_1 d_2 \cdots d_{k-1}}$. *Then*

$$\lceil r \rceil \leq \mathrm{mlt}_b = \mathrm{mlt}_e \leq \mathrm{mlt}_u \leq \lceil r \rceil + 1.$$

We note that for the case $k = 2$, a complete answer is known [12]. The case $k = 1$ is trivial. Corollary 1.3 gives nearly tight bounds on sample size thresholds. However, we note that for any particular choice of $d_1, \ldots, d_k$, we can always use Theorem 1.1 to get exact sample size thresholds.

Example 1.4. Consider the tensor normal model for 3-tensors (i.e., $k = 3$) with local dimensions $(d_1, d_2, d_3) = (3, 4, 5)$, which was used for simulation studies in [27]. Our Corollary 1.3 asserts that

$$1 \leq \mathrm{mlt}_b = \mathrm{mlt}_e \leq \mathrm{mlt}_u \leq 2.$$

To determine the precise thresholds, we compute $R = \Delta = 60m - 48$ and $g_{\max} = 1$. For $m = 1$, we have $R > 0$ and $\Delta \geq -1$. Thus part (1) of Theorem 1.1 shows that already for a single sample the MLE almost surely exists uniquely. We conclude that $\mathrm{mlt}_b = \mathrm{mlt}_e = \mathrm{mlt}_u = 1$.

The preceding example also shows that the criterion in Equation (15) of [27] (which in the notation of Corollary 1.3 claims that $\mathrm{mlt}_e = \lceil r \rceil + 1$) is in general not tight: that is, it only holds with an inequality.

1.3. Main results in invariant theory

Recently, Améndola, Kohn, Reichenbach and Seigal [2] established a connection between a class of Gaussian models called Gaussian group models and the invariant theory of a corresponding group action (see Theorem 2.3). We revisit this connection in Section 2. As mentioned previously, the group action that corresponds to tensor normal models is the tensor action. Given natural numbers $d_1, \ldots, d_k, m \in \mathbb{Z}_{>0}$, we denote by $\rho_{d_1, \ldots, d_k; m}$ the natural representation of $G = \mathrm{SL}_{d_1}(\mathbb{F}) \times \cdots \times \mathrm{SL}_{d_k}(\mathbb{F})$ on $V = \mathbb{F}^{d_1, \ldots, d_k; m}$. Theorem 1.1 is a consequence of the following invariant-theoretic result (see Section 2 for the definitions of stability).

Theorem 1.5. *Let* $\mathbb{F} = \mathbb{R}$ *or* $\mathbb{C}$. *Consider the tensor representation* $\rho = \rho_{d_1, \cdots, d_k; m}$. *Let* $R = R(d_1, \ldots, d_k; m)$, $\Delta = \Delta(d_1, \ldots, d_k; m)$ *and* $g_{\max} = g_{\max}(d_1, \ldots, d_k)$. *Then*

1. *If* $R > 0$, *then* ρ *is generically polystable. Furthermore:*
   - *If* $m \geq 2$, *then* $R \geq g_{\max}^2$, *and* ρ *is generically stable if and only if* $R > g_{\max}^2$ *or* $g_{\max} = 1$.
   - *If* $m = 1$, *then* $\Delta \geq -2$, *and* ρ *is generically stable if and only if* $\Delta \geq -1$.
2. *If* $R = 0$, *then* ρ *is generically polystable. It is generically stable if and only if* $g_{\max} = 1$.
3. *If* $R < 0$, *then* ρ *is unstable.*

While the preceding theorem gives a nice and uniform characterisation, it is essentially a reformulation of the following result, which is *recursive* in nature, but more enlightening.

Theorem 1.6. *Let* $\mathbb{F} = \mathbb{R}$ *or* $\mathbb{C}$. *Consider the tensor representation* $\rho = \rho_{d_1, \cdots, d_k; m}$. *Without loss of generality, assume* $d_1 \leq d_2 \leq \cdots \leq d_k$. *Then*

1. *If* $d_k > d_1 \cdots d_{k-1} m$, *then* ρ *is not generically semistable.*
2. *If* $d_k = d_1 \cdots d_{k-1} m$, *then* ρ *is generically polystable. It is generically stable if and only if* $d_1 = \cdots = d_{k-1} = 1$.
3. *If* $\frac{d_1 \cdots d_{k-1} m}{2} < d_k < d_1 \cdots d_{k-1} m$, *then* ρ *is generically semistable (polystable, stable) if and only if the same is true if we replace* d_k *by* $d'_k = d_1 \cdots d_{k-1} m - d_k$. *Note that* $1 \leq d'_k < d_k$.
4. *If* $d_k \leq \frac{d_1 \cdots d_{k-1} m}{2}$, *then* ρ *is generically polystable. Further, it is not generically stable if and only if* $(d_1, \ldots, d_k; m) = (1, \ldots, 1, 2, d, d; 1)$ *or* $(1, \ldots, 1, 1, d, d; 2)$ *for some* $d \geq 2$.

Moreover, if ρ *is not generically semistable, then it is unstable.*

Part (3) of Theorem 1.6 above is a reflection of the fact that the property of being generically semistable (polystable, stable) is unchanged under an operation known as a *castling transform*. Castling transforms played a crucial role in Sato and Kimura's classification of prehomogeneous vector spaces [35] (see also [39]). Its origins can be traced back to at least Elashvili's paper [15].

As a corollary of Theorems 1.5 and 1.6, we can derive a formula for the dimension of the GIT quotient (see Section 7 for a definition) of $V = \mathbb{F}^{d_1,\ldots,d_k;m}$ for the action of $G = \mathrm{SL}_{d_1} \times \cdots \times \mathrm{SL}_{d_k}$. This generalises the result of [4], where the dimension was computed in the case that $m = 1$.

Theorem 1.7. *Let* $\mathbb{F} = \mathbb{C}$. *Consider the natural action of* $G = \mathrm{SL}_{d_1} \times \cdots \times \mathrm{SL}_{d_k}$ *on* $V = \mathbb{F}^{d_1,\ldots,d_k;m}$. *Let* δ *denote the dimension of the GIT quotient* $\mathbb{P}V /\!\!/ G$.

1. *If $R < 0$, then the GIT quotient is empty.*
2. *If $R = 0$, then $\delta = 0$. In fact, the GIT quotient is a single point.*
3. *If $R > 0$, then*

$$\delta = \begin{cases} \max(g_{\max} - 3, 0) & \text{if } m = 1 \text{ and } \Delta = -2, \\ g_{\max} & \text{if } m = 2 \text{ and } R = g_{\max}^2 > 1, \\ \Delta & \text{otherwise.} \end{cases}$$

Organisation of the paper

In Section 2, we revisit the general connection between Gaussian group models and invariant theory and discuss the relevant notions of stability. In Section 3, we introduce castling transforms and discuss how they preserve stability. In Section 4, this is used as the key ingredient to derive our recursive characterisation (Theorem 1.6). In Section 5, we deduce our uniform characterisation (Theorem 1.5) from the former. In Section 6, we prove our main results on sample size thresholds for tensor normal models (Theorem 1.1 and Corollary 1.3). Finally, in Section 7, we compute the dimension of the GIT quotient (Theorem 1.7).

2. Gaussian group models and invariant theory

In this section, we first discuss the general setup of invariant theory. Then we define Gaussian group models and their connection to notions of generic stability in invariant theory. Finally, we discuss some general criteria from the literature useful for characterising generic stability.

Let $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. Let G be a group. A *representation* of G is an action of G on a (finite-dimensional) vector space V (over $\mathbb{F}$) by linear transformations. This is captured succinctly as a group homomorphism $\rho \colon G \to \mathrm{GL}(V)$. In particular, an element $g \in G$ acts on V by the linear transformation $\rho(g)$. We write $g \cdot v$ or gv to mean $\rho(g)v$. The G-orbit of $v \in V$ is the set of all vectors that you can get from v by applying elements of the group: that is,

$$O_v := \{gv \mid g \in G\} \subseteq V.$$

Throughout this paper, we will only consider the setting where G is a linear algebraic group (over $\mathbb{F}$) and where the action is rational: that is, $\rho \colon G \to \mathrm{GL}(V)$ is a morphism of algebraic groups.

We denote by $\mathbb{F}[V]$ the ring of polynomial functions on V (also known as the coordinate ring of V). A polynomial function $f \in \mathbb{F}[V]$ is called *invariant* if $f(gv) = f(v)$ for all $g \in G$ and $v \in V$. In other words, a polynomial is called invariant if it is constant on orbits. The invariant ring is

$$\mathbb{F}[V]^G := \{f \in \mathbb{F}[V] \mid f(gv) = f(v) \ \forall \, g \in G, v \in V\}.$$

The invariant ring has a natural grading by degree: that is, $\mathbb{F}[V]^G = \oplus_{d=0}^\infty \mathbb{F}[V]_d^G$, where $\mathbb{F}[V]_d^G$ consists of all invariant polynomials that are homogeneous of degree d. For $v \in V$, we define the stabiliser subgroup $G_v := \{g \in G \mid gv = v\}$, and we denote by $\overline{O}_v$ the closure of the orbit O_v.

Remark 2.1. To define the closure, we need to specify a topology on *V*. In this paper, we only use the fields $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. Hence, unless otherwise specified, we will use the standard Euclidean topology on *V* for orbit closures. At times we will also need to use the Zariski topology, but we will be careful in specifying it each time. For $\mathbb{F} = \mathbb{C}$, the orbit closure with regard to the Euclidean topology agrees with the orbit closure with regard to the Zariski topology (in the setting of rational actions of reductive groups).

We make a few definitions, the significance of which will become clear in the following subsections.

Definition 2.2. Let $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, and let *G* be an algebraic group (over $\mathbb{F}$) with a rational action on a vector space *V* (over $\mathbb{F}$), given by $\rho \colon G \to \mathrm{GL}(V)$. Let *K* denote the kernel of the homomorphism ρ. Give *V* the standard Euclidean topology. Then for $v \in V$, we say *v* is

○ *unstable* if $0 \in \overline{O}_v$;
○ *semistable* if $0 \notin \overline{O}_v$;
○ *polystable* if $v \neq 0$ and O_v is closed;
○ *stable* if *v* is polystable and the quotient G_v / K is finite.

2.1. *Gaussian group models*

For a subgroup $G \subseteq \mathrm{GL}_n$, we define an associated *Gaussian group model* by the following family of concentration matrices:

$$\mathcal{M}_G := \{g^\dagger g \mid g \in G\},$$

where $g^\dagger = \bar{g}^T$ denotes the adjoint. So for a concentration matrix $\Psi = g^\dagger g \in \mathcal{M}_G$ and an *m*-tuple of samples $Y = (Y_1, \ldots, Y_m) \in (\mathbb{F}^n)^m$, the log-likelihood function in equation (1.1) simplifies to

$$l_Y(\Psi) = \frac{m}{2} \log(\det(g^\dagger g)) - \frac{1}{2}\|g \cdot Y\|^2,$$

where $\|\cdot\|$ denotes the ℓ_2-norm on $(\mathbb{F}^n)^m \cong \mathbb{F}^{nm}$, and we note that *G* acts on $(\mathbb{F}^n)^m$ by the diagonal action $g \cdot Y = (gY_1, \ldots, gY_m)$.

The following result was proved in [2, Theorems 6.10 and 6.24]. It connects maximum likelihood estimation in Gaussian group models to the stability notions introduced in Definition 2.2.

Theorem 2.3 [2]. *Let $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. Let $G \subseteq \mathrm{GL}_n$ be a Zariski-closed subgroup that is closed under adjoints and nonzero scalar multiples. Let $G_{\mathrm{SL}} = \{g \in G \mid \det(g) = 1\} \subseteq G$, and let $Y \in (\mathbb{F}^n)^m$ be an m-tuple of samples. Then for the diagonal action of G_{SL}, we have*

○ *Y is semistable $\Longleftrightarrow$ l_Y is bounded from above;*
○ *Y is polystable $\Longleftrightarrow$ an MLE exists (i.e., l_Y has a maximum);*
○ *Y is stable $\Longrightarrow$ there exists a unique MLE (i.e., l_Y has a unique maximum). If $\mathbb{F} = \mathbb{C}$, the converse also holds: that is, there exists a unique MLE $\Longrightarrow$ Y is stable.*

Moreover, if Ψ is an MLE given Y, then the set of all MLEs given Y is $\{g^\dagger \Psi g \mid g \in (G_{\mathrm{SL}})_Y\}$.

Remark 2.4. In the setting of the above theorem, for $h \in G_{\mathrm{SL}}$, we also have

$$\{\text{MLEs given } h \cdot Y\} = (h^{-1})^\dagger \{\text{MLEs given } Y\} h^{-1}.$$

Thus, for any $h \in G_{\mathrm{SL}}$, the MLE given *Y* is unique if and only if the MLE given $h \cdot Y$ is unique.

Now let $\mathbb{F} = \mathbb{R}$, and suppose *Y* is already a point with minimal norm in its orbit. Then for an appropriate $\lambda \in \mathbb{R}_{>0}$, we have that λI is an MLE and the set of all MLEs is $\{\lambda g^\dagger g \mid g \in (G_{\mathrm{SL}})_Y\}$. In particular, we have a unique MLE if and only if $(G_{\mathrm{SL}})_Y \subseteq O_n$, the orthogonal group. Further, since $(G_{\mathrm{SL}})_Y$ is closed, it must be compact. The stabiliser of any other point in its G_{SL}-orbit is obtained by

conjugation and remains compact. In particular, if Y is any tuple of samples such that the MLE exists uniquely, then $(G_{SL})_Y$ is compact. This will be important to us, so we record the statement for later use:

Corollary 2.5. *Suppose we are in the setting of Theorem 2.3, with $\mathbb{F} = \mathbb{R}$. If the MLE given Y exists uniquely, then $(G_{SL})_Y$ is compact.*

When $\mathbb{F} = \mathbb{C}$, the same hypothesis and argument shows that $(G_{SL})_Y$ is finite. However, we will only need Corollary 2.5 in the case that $\mathbb{F} = \mathbb{R}$.

2.2. Notions of generic stability

Let G be an algebraic group (over $\mathbb{F}$), and let V be a rational representation (over $\mathbb{F}$). Then we define

$$V^{\mathrm{ss}} = \{v \in V \mid v \text{ is } G\text{-semistable}\},$$
$$V^{\mathrm{ps}} = \{v \in V \mid v \text{ is } G\text{-polystable}\},$$
$$V^{\mathrm{st}} = \{v \in V \mid v \text{ is } G\text{-stable}\}.$$

We call V^{ss} (respectively $V^{\mathrm{ps}}, V^{\mathrm{st}}$) the *semistable* (respectively *polystable*, *stable*) *locus*. If the group is not clear from context, then we write $V^{G\text{-ss}}, V^{G\text{-ps}}, V^{G\text{-st}}$.

Definition 2.6. Let $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, and let G be an algebraic group (over $\mathbb{F}$) with a rational action on a vector space V (over $\mathbb{F}$). Then we say V is *generically G-semistable* (respectively *polystable*, *stable*) if V^{ss} (respectively $V^{\mathrm{ps}}, V^{\mathrm{st}}$) contains a nonempty Zariski-open subset of V. Further, we say that V is *unstable* if $V^{\mathrm{ss}} = \emptyset$.

These notions are particularly well-behaved in the case that $\mathbb{F} = \mathbb{C}$, as we will see in the following. We refer to [12, Corollary 2.15, Lemma 2.16] for a succinct proof of the following standard result:

Lemma 2.7. *Suppose $\mathbb{F} = \mathbb{C}$. Let V be a rational representation of a complex reductive group G. Then the subsets V^{ss} and V^{st} are Zariski-open and the subset V^{ps} is Zariski-constructible: that is, it is a union of Zariski-locally closed subsets. Moreover, V is generically semistable if and only if it is not unstable.*

Zariski-open subsets of a vector space, whenever nonempty, are complements of lower-dimensional subvarieties, which have Lebesgue measure zero. On the other hand, Zariski-constructible subsets of a vector space have Lebesgue measure zero unless they contain a Zariski-open subset, in which case their complement has Lebesgue measure zero. Hence, we can conclude the following:

Corollary 2.8. *Suppose we are in the setting of Theorem 2.3, with $\mathbb{F} = \mathbb{C}$. Fix a number of samples m, and let $V = (\mathbb{C}^n)^m$. Then for the diagonal action of G_{SL}, we have*

○ *V is generically semistable $\Longleftrightarrow l_Y$ is almost surely bounded from above;*
○ *V is generically polystable $\Longleftrightarrow$ an MLE exists almost surely;*
○ *V is generically stable $\Longleftrightarrow$ there exists a unique MLE almost surely;*
○ *V is unstable $\Longleftrightarrow l_Y$ is always unbounded from above.*

Moreover, the first and last conditions are complementary. Here we say a property holds almost surely if it holds for all Y in V up to a set of Lebesgue measure zero.

Let us also mention one lemma that will be useful for us later:

Lemma 2.9. *Suppose G is a complex algebraic group, and let V be a rational representation over $\mathbb{C}$. If $V^{\oplus m}$ is generically G-stable (respectively G-semistable), then $V^{\oplus n}$ is generically G-stable (respectively G-semistable) for all $n \geq m$ with respect to the diagonal actions of G.*

Proof. Suppose $V^{\oplus m}$ is generically G-stable. We have an inclusion $(V^{\oplus m})^{\mathrm{st}} \subseteq (V^{\oplus n})^{\mathrm{st}}$ with respect to the diagonal actions of G. So $(V^{\oplus n})^{\mathrm{st}}$ is nonempty, and further it is Zariski open by Lemma 2.7. Thus, $V^{\oplus n}$ is generically G-stable. The argument for semistability is similar.   □

2.3. Stabilisers in general position

Let $\mathbb{F} = \mathbb{C}$ for this section. Let V be a rational representation of a reductive group G. We say that H is a *stabiliser in general position (s.g.p.)* if there is a nonempty Zariski-open subset $U \subseteq V$ such that for all $v \in U$, the stabiliser G_v is isomorphic to H. The s.g.p. is unique up to conjugation. Its existence is far from obvious and follows from Luna's slice theorem; see, for example, [38, Theorem 7.2]. Indeed, when $\mathbb{F} = \mathbb{R}$, stabilisers in general position often do not exist.

Matsushima's criterion tells us that if an orbit of a point is closed, then the stabiliser is reductive. Hence, if V is generically polystable, then the s.g.p. must be reductive. The converse was proved by Popov:

Theorem 2.10 [30]. *Let $\rho\colon G \to \mathrm{GL}(V)$ be a rational representation of a reductive group. Then V is generically polystable if and only if the stabiliser in general position is reductive.*

Corollary 2.11. *Let $\rho\colon G \to \mathrm{GL}(V)$ be a rational representation of a reductive group, and let K denote the kernel of ρ. Let H be the stabiliser in general position. The following are equivalent:*

1. *V is generically stable;*
2. *$\dim(H) = \dim(K)$;*
3. *$\dim(G_v) = \dim(K)$ for some $v \in V$.*

Proof. Clearly (1) $\implies$ (2) $\implies$ (3). For (2) $\implies$ (1), observe that $\dim(H) = \dim(K)$ implies that G_v/K is finite for generic $v \in V$. The kernel of a morphism of (affine) algebraic groups between reductive groups is reductive, so K is reductive. Since G_v/K is finite (for generic $v \in V$), this means G_v and K have the same identity component and hence G_v is also reductive. In particular, it means H is reductive. Hence V is generically polystable by Theorem 2.10 and, further, generically stable because G_v/K is finite for generic $v \in V$.

For (3) $\implies$ (2), we observe that the set of points $U = \{v \in V \mid \dim(G_v) \leq \dim(K)\}$ is Zariski open. Note that $U = \{v \in V \mid \dim(G_v) = \dim(K)\}$ since $K \subseteq G_v$ for all $v \in V$. Since U is nonempty Zariski open, it follows that $\dim(H) = \dim(K)$ as well.   □

2.4. A criterion for generic (poly)stability

Let still be $\mathbb{F} = \mathbb{C}$ for this section. Since the late 1960s, there has been an interest in classifying actions that are generically polystable or stable; see, for example, [3, 15, 35, 29]. From this line of research, we will recall a few results that will be important for us.

If S is a simple algebraic group, then the Killing form defined by $(X, Y) \mapsto \mathrm{tr}(\mathrm{ad}(X)\mathrm{ad}(Y))$ is a nondegenerate symmetric S-invariant bilinear form on the Lie algebra $\mathfrak{s}$ of S. Up to a scalar, $\mathfrak{s}$ has only one S-invariant symmetric bilinear form. If $\rho\colon S \to \mathrm{GL}(V)$ and $d\rho\colon \mathfrak{s} \to \mathrm{End}(V)$ is the corresponding representation of the Lie algebra, then $(X, Y) \mapsto \mathrm{tr}(d\rho(X)d\rho(Y))$ is a nonzero symmetric S-invariant bilinear form on $\mathfrak{s}$. So there is a constant $\iota_S(V)$, called the *index* of the representation, such that

$$\mathrm{tr}(d\rho(X)d\rho(Y)) = \iota_S(V)\,\mathrm{tr}(\mathrm{ad}(X)\mathrm{ad}(Y))$$

for all $X, Y \in \mathfrak{s}$. The index is additive. Furthermore, we have $\iota_{\mathrm{SL}_n}(\mathbb{C}^n) = \frac{1}{2n}$ for the defining representation of SL_n.

Andreev, Vinberg and Elashvili proved the following criterion for generic stability in [3, Theorem].

Theorem 2.12 [3]. *Let $\rho\colon G \to \mathrm{GL}(V)$ be a rational representation of a connected semisimple[1] group. Let H be the stabiliser in general position. If $\iota_S(V) > 1$ for all simple normal subgroups $S \subseteq G$, then $\dim(H) = 0$. In particular, V is generically G-stable.*

Elashvili proved a very similar criterion for generic polystability in [15, Theorem 2].

---

[1]Semisimple groups are reductive.

Theorem 2.13 [15]. *Let* $\rho \colon G \to \mathrm{GL}(V)$ *be a rational representation of a connected semisimple group. Let H be the stabiliser in general position. If* $\iota_S(V) \geq 1$ *for all simple normal subgroups* $S \subseteq G$, *then the Lie algebra of H is the Lie algebra of a torus. In particular, H is reductive, so V is generically G-polystable.*

Just to put these results in context, let us consider the tensor action: that is, the action of $G = \prod_{i=1}^{k} \mathrm{SL}_{d_k}$ on $\mathbb{C}^{d_1,\ldots,d_k;m}$. In this case, G is a connected semisimple group, and its simple normal subgroups are just $\mathrm{SL}_{d_1}, \mathrm{SL}_{d_2}, \ldots, \mathrm{SL}_{d_k}$ and the index for each SL_{d_i} is $\frac{m \prod_{j \neq i} d_j}{2 d_i}$.

Finally, Elashvili has classified all irreducible representations that satisfy the hypotheses of Theorem 2.13 but are not generically stable; see [15, Theorem 9] and Theorem 4.3 below.

3. Castling transforms

In this section, we take $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. Let $\rho \colon G \to \mathrm{GL}(V)$ be an n-dimensional representation of an algebraic group G. We will assume $\rho(G) \subseteq \mathrm{SL}(V)$. For $0 < k < n$, we have a natural action of $G \times \mathrm{SL}_k$ on $V \otimes \mathbb{F}^k$, where G acts on V and SL_k acts on $\mathbb{F}^k$. Similarly, we have an action of G on V^* and of SL_{n-k} on $\mathbb{F}^{n-k}$, which together gives an action of $G \times \mathrm{SL}_{n-k}$ on $V^* \otimes \mathbb{F}^{n-k}$. We refer to the action of $G \times \mathrm{SL}_{n-k}$ on $V^* \otimes \mathbb{F}^{n-k}$ as a *castling transform* of the action of $G \times \mathrm{SL}_k$ on $V \otimes \mathbb{F}^k$.

The main feature of castling transforms is that we get a bijection between the $G \times \mathrm{SL}_k$-orbits in a nonempty Zariski-open subset of $V \otimes \mathbb{F}^k$ and the $G \times \mathrm{SL}_{n-k}$-orbits in a nonempty Zariski-open subset of $V^* \otimes \mathbb{F}^{n-k}$. Moreover, this bijection of orbits preserves stabilisers up to isomorphism. Hence, when $\mathbb{F} = \mathbb{C}$, the stabiliser in general position is preserved under castling transforms. Moreover, generic semistability/polystability/stability will also be preserved under castling transforms. We will now explain all this in more detail, but first we need to recall Grassmannians.

3.1. Grassmannians

Let $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. Suppose V is an n-dimensional vector space over $\mathbb{F}$. Let $\mathrm{Gr}(k, V)$ denote the Grassmannian of k-planes in V. It is naturally embedded in $\mathbb{P}(\bigwedge^k(V))$ as a closed subvariety cut out by the Plücker relations, where $\bigwedge^k(V)$ denotes the kth exterior power of V.

This embedding is constructed as follows. Identify V with $\mathbb{F}^n$ by choosing a basis $e_1, \ldots, e_n$. Then a basis for $\bigwedge^k(V)$ is $\{e_{i_1} \wedge e_{i_2} \wedge \cdots \wedge e_{i_k} \mid 1 \leq i_1 < i_2 < \cdots < i_k \leq n\}$. For any subset $I \subseteq [n]$ of size k, we write e_I to denote $e_{i_1} \wedge e_{i_2} \wedge \cdots \wedge e_{i_k}$, where $I = \{i_1, \ldots, i_k\}$ with the i_js in increasing order. We write Δ_I to denote the coordinate corresponding to e_I. Now, for any subspace L of V of dimension k, take independent vectors $l_1, \ldots, l_k$ in L and consider the point $[l_1 \wedge l_2 \wedge \cdots \wedge l_k] \in \mathbb{P}(\bigwedge^k(V))$. This point is independent of the choice of l_i and only depends on the subspace L. Thus, we obtain an injective map $\mathrm{Gr}(k, V) \to \mathbb{P}(\bigwedge^k(V))$ whose image is a closed subvariety. This map is called the Plücker embedding and endows the Grassmannian with the structure of a projective variety. We refer to [17, 44, 31] for more details on Grassmannians.

The affine cone over the Grassmannian $\widehat{\mathrm{Gr}}(k, V)$ is a closed subvariety of $\bigwedge^k(V)$. Note that $\widehat{\mathrm{Gr}}(k, V) = \{v_1 \wedge v_2 \wedge \cdots \wedge v_k \mid v_i \in V\}$. If the v_is are linearly dependent, then $v_1 \wedge v_2 \wedge \cdots \wedge v_k = 0$; otherwise it is nonzero. Let $\{e_1, \ldots, e_k\}$ denote the standard basis for $\mathbb{F}^k$, and define

$$U = \left\{ \textstyle\sum_{i=1}^{k} v_i \otimes e_i \in V \otimes \mathbb{F}^k \mid v_1, \ldots, v_k \text{ are linearly independent} \right\}. \tag{3.1}$$

Then we have a map

$$\pi = \pi_{k,V} \colon U \longrightarrow \widehat{\mathrm{Gr}}(k, V) \setminus \{0\}, \qquad \textstyle\sum_{i=1}^{k} v_i \otimes e_i \longmapsto v_1 \wedge v_2 \wedge \cdots \wedge v_k.$$

We claim that U is a Zariski-locally trivial principal SL_k-bundle over $\widehat{\mathrm{Gr}}(k, V) \setminus \{0\}$. It is straightforward to see that it is a principal SL_k-bundle, because $v_1 \wedge v_2 \wedge \cdots \wedge v_k = w_1 \wedge w_2 \wedge \cdots \wedge w_k$ if and only if there is a matrix $A = (a_{ij}) \in \mathrm{SL}_k$ such that $\sum_i a_{ij} v_j = w_i$ for all i. To see that it is Zariski-locally trivial

needs an explanation. A similar result, namely that U is a Zariski-locally trivial principal GL_k-bundle over $\mathrm{Gr}(k, V)$, is well known; see, for example, [31, pg. 511]. We modify their argument appropriately.

First, we note that $\widehat{\mathrm{Gr}}(k, V) \setminus \{0\}$ is covered by affine open subsets $\{X_I : I \subseteq [n], |I| = k\}$, where $X_I := \{p \mid \Delta_I(p) \neq 0\}$. If we identify V with $\mathbb{F}^n$ as mentioned above, U can be viewed as the $k \times n$ matrices of full rank. For a matrix $M \in \mathrm{Mat}_{k,n}$ and a subset $I \subseteq [n]$ of size k, let M_I denote the $k \times k$ submatrix of M obtained by considering the columns labeled by elements in I, and let $p_I(M) = \det(M_I)$. Then $\pi^{-1}(X_I) = \{M \in \mathrm{Mat}_{k,n} \mid p_I(M) \neq 0\}$. Without loss of generality, we can take $I = \{1, 2, \dots, k\}$, so we have an isomorphism $\pi^{-1}(X_I) \rightarrow \mathrm{Mat}_{k,n-k} \times \mathbb{F}^* \times \mathrm{SL}_k$ given by $M = [A \mid B] \mapsto (DA^{-1}B, \det(A), AD^{-1})$, where D is the diagonal matrix with diagonal entries $(\det(A), 1, 1, \dots, 1)$. The map in the reverse direction is $(P, \lambda, Q) \mapsto [QD \mid QP]$, where $D = \mathrm{diag}(\lambda, 1, \dots, 1)$. Now, observing that $X_I \cong \mathrm{Mat}_{k,n-k} \times \mathbb{F}^{*}$ [2] gives us an isomorphism $\pi^{-1}(X_I) \longrightarrow X_I \times \mathrm{SL}_k$.

Everything we said above also works if you consider the Euclidean topology because Zariski-open subsets are open in the Euclidean topology and polynomial maps are continuous in the Euclidean topology as well. Hence, U is also a locally trivial principal SL_k-bundle over $\widehat{\mathrm{Gr}}(k, V) \setminus \{0\}$ in the Euclidean topology.

The projection of a locally trivial bundle onto its base is an open map. One can check this condition on a trivialising cover of the base. In other words, it suffices to check that projection of a trivial bundle onto its base is open. For the Euclidean topology, it is well known that projection maps are open. For the Zariski topology, projection maps are also open. When the underlying field is algebraically closed, this follows from flatness, but the statement remains true even when the underlying field is not algebraically closed:

Lemma 3.1. *Suppose X and Y are affine $\mathbb{F}$-varieties; then the projection map $\pi \colon X \times Y \rightarrow X$ is an open map in the Zariski topology.*

Proof. We have $X = \mathbb{V}(f_1, \dots, f_r)$ and $Y = \mathbb{V}(g_1, \dots, g_s)$, where $\mathbb{V}(\dots)$ denotes the zero locus of a collection of polynomials. Now, suppose $U = \mathbb{V}(p_1, \dots, p_t)^c$ is a Zariski-open subset of $X \times Y$. Then

$$
\begin{aligned}
\pi(U) &= \{a \mid \exists b \in Y : (a, b) \in U\} \\
&= \{a \mid \exists b \in Y, 1 \leq i \leq t : p_i(a, b) \neq 0\} \\
&= \mathbb{V}(\{p_{i,b}\}_{b \in Y, 1 \leq i \leq t})^c,
\end{aligned}
$$

where $p_{i,b} = p_i(-, b)$. Thus $\pi(U)$ is Zariski-open. Note that even though $p_{i,b}$ is an infinite collection of polynomials, one can extract a finite subset with the same zero locus by the Hilbert basis theorem. $\square$

To summarise, we get the following result:

Lemma 3.2. *Let V, U and $\pi_{k,V}$ be defined as above. Then U is a Zariski-locally trivial principal SL_k-bundle over $\widehat{\mathrm{Gr}}(k, V) \setminus \{0\}$ via the map $\pi_{k,V}$. In particular, $\pi_{k,V}$ is an open map (and also a quotient map) when considering either the Zariski or Euclidean topology.*

3.2. Castling transforms

Let $\rho \colon G \rightarrow \mathrm{GL}(V)$ be a representation of an algebraic group G, and we will assume $\rho(G) \subseteq \mathrm{SL}(V)$. Let $\dim(V) = n$. We have an action of $G \times \mathrm{SL}_k$ on $V \otimes \mathbb{F}^k$ and an action of $G \times \mathrm{SL}_{n-k}$ on $V^* \otimes \mathbb{F}^{n-k}$. Let

$$
U = \left\{ \textstyle\sum_{i=1}^{k} v_i \otimes e_i \in V \otimes \mathbb{F}^k \mid v_1, \dots, v_k \text{ are linearly independent} \right\} \subseteq V \otimes \mathbb{F}^k,
$$

as in equation (3.1), and let

$$
U' = \left\{ \textstyle\sum_{i=1}^{n-k} w_i \otimes e_i \in V^* \otimes \mathbb{F}^{n-k} \mid w_1, \dots, w_{n-k} \text{ are linearly independent} \right\} \subseteq V^* \otimes \mathbb{F}^{n-k}.
$$

---

[2]It is well known in the projective setting that the locus where $\Delta_I(p) \neq 0$ is isomorphic to $\mathrm{Mat}_{k,n-k}$, and we are just pulling back to the affine cone.

Since U is a principal SL_k-bundle over $\widehat{\mathrm{Gr}}(k, V) \setminus \{0\}$, we have a bijection between the SL_k-orbits in U and the points of $\widehat{\mathrm{Gr}}(k, V) \setminus \{0\}$. This bijection is G-equivariant since $\pi_{k,V}$ is G-equivariant and the actions of G and of SL_k on $V \otimes \mathbb{C}^k$ commute. So, we have G-equivariant bijections:

$$\mathrm{SL}_k\text{-orbits in } U \;\longleftrightarrow\; \widehat{\mathrm{Gr}}(k, V) \setminus \{0\} \;\longleftrightarrow\; \widehat{\mathrm{Gr}}(n - k, V^*) \setminus \{0\} \;\longleftrightarrow\; \mathrm{SL}_{n-k}\text{-orbits in } U'. \quad (3.2)$$

The first bijection was explained above, and the last bijection follows by the same argument. The middle bijection comes from the well understood $\mathrm{SL}(V)$-equivariant isomorphism $\bigwedge^k(V) \cong \bigwedge^{n-k}(V^*)$. The following result is implicit in [15], but we furnish a proof for completeness.

Lemma 3.3. *Let $T \in U$. Then we have an isomorphism of algebraic groups*

$$\mathrm{Stab}_G(\pi_{k,V}(T)) \cong \mathrm{Stab}_{G \times \mathrm{SL}_k}(T).$$

Proof. This holds since $\pi_{k,V}$ is a G-equivariant principal SL_k-bundle. Indeed, let $p \colon G \times \mathrm{SL}_k \to G$ denote the projection onto the first factor. It is easy to see that $p(\mathrm{Stab}_{G \times \mathrm{SL}_k}(T)) \subseteq \mathrm{Stab}_G(\pi_{k,V}(T))$. Now suppose $g \in \mathrm{Stab}_G(\pi_{k,V}(T))$. Then $\pi_{k,V}(T) = g \cdot \pi_{k,V}(T)$ implies that $\pi_{k,V}(T) = \pi_{k,V}(g \cdot T)$ by G-equivariance. Since $\pi_{k,V}$ is a principal SL_k-bundle, it follows that there exists a unique $A \in \mathrm{SL}_k$ such that $A \cdot (g \cdot T) = T$: that is, $(g, A) \cdot T = T$. Thus we have proved that every $g \in \mathrm{Stab}_G(\pi_{k,V}(T))$ has a unique preimage under p in $\mathrm{Stab}_{G \times \mathrm{SL}_k}(T)$. We conclude that p restricted to $\mathrm{Stab}_{G \times \mathrm{SL}_k}(T)$ is a (group) isomorphism onto its image, which is $\mathrm{Stab}_G(\pi_{k,V}(T))$.

To establish that this is an isomorphism of algebraic groups (over $\mathbb{F}$), we need to establish that it is an isomorphism of varieties. To do so, we give a map in the reverse direction as follows. Write $T = \sum_{i=1}^k v_i \otimes e_i$. Let $g \in \mathrm{Stab}_G(\pi_{k,V}(T))$. Since g stabilises the span of $v_1, \ldots, v_k$, we get that $g \cdot v_i = \sum_j c_{i,j}(g) \, v_j$, where the $c_{i,j}(g)$ are regular functions on $\mathrm{Stab}_G(\pi_{k,V}(T))$. Moreover, the matrix $C = (c_{i,j}(g))_{1 \le i, j \le k}$ is invertible. Then (g, C^{-1}) is the unique preimage of g in $\mathrm{Stab}_{G \times \mathrm{SL}_k}(T)$ under p. Thus the map $g \mapsto (g, C^{-1})$ is the inverse of p restricted to $\mathrm{Stab}_{G \times \mathrm{SL}_k}(T)$, and it is clearly a morphism of algebraic varieties. $\square$

As a consequence of the bijections in equation (3.2) and Lemma 3.3, we thus obtain the following corollaries.

Corollary 3.4. *We have a natural bijection between the $G \times \mathrm{SL}_k$-orbits in U and the $G \times \mathrm{SL}_{n-k}$ orbits in U' that preserves stabilisers (up to isomorphism).*

Corollary 3.5. *Let $\mathbb{F} = \mathbb{C}$. Then the stabiliser in general position for the action of $G \times \mathrm{SL}_k$ on $V \otimes \mathbb{C}^k$ is isomorphic to the stabiliser in general position for the action of $G \times \mathrm{SL}_{n-k}$ on $V^* \otimes \mathbb{C}^{n-k}$.*

In fact, the invariant ring is also preserved by castling transforms [35] (see also [21, Prop. 2.1]).

Lemma 3.6 [35]. *Let $\mathbb{F} = \mathbb{C}$. Then the invariant ring for the action of $G \times \mathrm{SL}_k$ on $V \otimes \mathbb{C}^k$ is (canonically) isomorphic to the invariant ring for the action of $G \times \mathrm{SL}_{n-k}$ on $V^* \otimes \mathbb{C}^{n-k}$.*

The discussion above culminates in the following result that will be very important for us:

Corollary 3.7. *Let $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. Then $V \otimes \mathbb{F}^k$ is generically $G \times \mathrm{SL}_k$-semistable (polystable, stable) if and only if $V^* \otimes \mathbb{F}^{n-k}$ is generically $G \times \mathrm{SL}_{n-k}$-semistable (polystable, stable).*

Proof. By [12, Proposition 2.23], it suffices to prove the statement for $\mathbb{F} = \mathbb{C}$. So, let us assume that $\mathbb{F} = \mathbb{C}$. Generic semistability is the same as having a nontrivial invariant ring. Hence, it follows from Lemma 3.6 that castling transforms preserve generic semistability. The fact that castling transforms preserve generic polystability follows from Corollary 3.5 and Theorem 2.10.

That castling transforms preserve generic stability follows similarly from Corollaries 3.5 and 2.11, provided we can show that the kernels of the two actions have the same dimension. To see this, let $K = \ker(\rho)$, where $\rho \colon G \to \mathrm{GL}(V)$ is the action of G on V. Now, let us consider the kernel of $\tilde{\rho} \colon G \times \mathrm{SL}_k \to \mathrm{GL}(V \otimes \mathbb{C}^k)$. For $(g, A) \in G \times \mathrm{SL}_k$, we have $\tilde{\rho}(g, A) = \rho(g) \otimes A$. So, if (g, A) is in the

kernel, then $\rho(g) = c\mathrm{I}$ and $A = c^{-1}I$ for some $c \in \mathbb{C}^*$. But $A \in \mathrm{SL}_k$, so c must be an kth root of unity. For each such c, the subvariety $H_c = \{g \in G \mid \rho(g) = cI\}$ is either empty or a coset of K. Since the kernel is a finite union of $H_c \times \{c^{-1}I\}$, its dimension equals the dimension of K. On the other hand, the kernel for the action of G on V^* is also K, so the same argument shows that the kernel for the action of $G \times \mathrm{SL}_{n-k}$ on $V^* \otimes \mathbb{C}^{n-k}$ also has the same dimension as K.   □

For complex Gaussian group models, we saw in Theorem 2.3 that invariant-theoretic stability notions characterise the boundedness of the log-likelihood function and the existence and uniqueness of MLEs precisely. However, for real models, the relation between generic stability and almost sure existence of a unique MLE is less tight. To bridge this gap, we will need the following results:

Lemma 3.8. *Suppose $P \subseteq V \otimes \mathbb{F}^k$ is a nonempty open subset in the Euclidean (respectively Zariski) topology. Then $(\pi_{n-k,V^*})^{-1}\pi_{k,V}(P \cap U)$ is a nonempty open subset of U' in the Euclidean (respectively Zariski) topology.*

Proof. Let us first argue this for Euclidean topology. Observe that $P \cap U$ is an open subset of $V \otimes \mathbb{C}^k$. Further, since U^c is a proper subvariety and hence has empty interior, we know that $P \cap U$ must be nonempty. Now the statement follows since $\pi_{k,V}$ is an open map by Lemma 3.2. The argument for Zariski topology is analogous.   □

An immediate corollary of the above lemma is the following:

Corollary 3.9. *Let $P = \{T \in V \otimes \mathbb{F}^k \mid \mathrm{Stab}_{G \times \mathrm{SL}_k}(T) \text{ is not compact}\}$. Similarly, let $P' = \{S \in V^* \otimes \mathbb{F}^{n-k} \mid \mathrm{Stab}_{G \times \mathrm{SL}_{n-k}}(T) \text{ is compact}\}$. Then P contains a nonempty Euclidean (respectively Zariski) open subset of $V \otimes \mathbb{F}^k$ if and only if P' contains a nonempty Euclidean (respectively Zariski) open subset of $V^* \otimes \mathbb{F}^k$.*

Proof. It suffices to prove one direction. Suppose P contains a nonempty Euclidean (respectively Zariski) open subset $\widetilde{P}$. Then by Lemma 3.8, $(\pi_{n-k,V^*})^{-1}\pi_{k,V}(\widetilde{P} \cap U)$ is a Euclidean (respectively Zariski) open subset of $V^* \otimes \mathbb{F}^{n-k}$, and it is contained in P' by Corollary 3.4.   □

We need to give a technical clarification in the above corollary with respect to the notion of compactness. There are two natural topologies one can give a Lie subgroup H of a Lie group G. The first is the inherent topology on H by virtue of being a Lie group in itself, and the second is the subspace topology by virtue of being a subspace of G. In the proof above, we are really using the inherent topology because the isomorphism of stabilisers furnished by Corollary 3.4 is an abstract isomorphism. However, we will later need to use the lemma in the context of Corollary 2.5, which refers to the subspace topology. While for immersed Lie subgroups, the inherent topology can differ from the subspace topology, the two topologies coincide for embedded Lie subgroups. Since stabiliser subgroups are closed, they are embedded Lie subgroups, and there is no ambiguity.

3.3. Castling transforms for tensor actions

We now discuss explicitly the relevance of castling transforms to tensor actions and hence to tensor normal models. Here we are interested in the action of $\prod_{i=1}^k \mathrm{SL}_{d_i}$ on $\mathbb{F}^{d_1,\dots,d_k;m}$, which we succinctly denote by $\rho_{d_1,\dots,d_k;m}$. The ground field $\mathbb{F}$ is assumed to be either $\mathbb{R}$ or $\mathbb{C}$. If we need to specify it, we will add a subscript.

Let $G = \prod_{i=1}^{k-1} \mathrm{SL}_{d_i}$, and consider its natural action on $V = \mathbb{F}^{d_1,\dots,d_{k-1};m}$, which in our notation is $\rho_{d_1,\dots,d_{k-1};m}$. Then the action of $G \times \mathrm{SL}_{d_k}$ on $V \otimes \mathbb{F}^{d_k}$ is simply $\rho_{d_1,\dots,d_k;m}$. It is well known that V and V^* are related by an automorphism on the group G, which does not affect any of the notions of stability. [3]Hence, we call $\rho_{d_1,\dots,N-d_k;m}$ the *castling transform* of $\rho_{d_1,\dots,d_k;m}$, where $N = \dim V = md_1 \cdots d_{k-1}$, and we assume that $N > d_k$. Thus Corollary 3.7 implies the following important result:

---

[3]If we compose a representation ρ of $\mathrm{SL}(d)$ with the automorphism $g \mapsto g^{-T}$, the result is isomorphic to the dual representation of ρ, and similarly for the product group G.

Corollary 3.10. *Let $d_1, \ldots, d_k, m \in \mathbb{Z}_{>0}$, and suppose that $N = m \prod_{i=1}^{k-1} d_i > d_k$. Then $\rho_{d_1,\ldots,d_k;m}$ is generically semistable (polystable, stable) if and only if $\rho_{d_1,\ldots,N-d_k;m}$ is generically semistable (polystable, stable).*

Given this result, we will make some definitions for later use. For positive integers $d_1, \ldots, d_k$ and m, we call $(d_1, \ldots, d_k; m)$ a *datum* and $\rho_{d_1,\ldots,d_k;m}$ the corresponding representation. Observe that permuting the d_i leaves the group and representation unchanged up to isomorphism and hence does not change the generic stability properties of the representation.

Definition 3.11. We say two data $(d_1, \ldots, d_k; m)$ and $(d'_1, \ldots, d'_k; m)$ are *castling-equivalent* if $\rho_{d_1,\ldots,d_k;m}$ and $\rho_{d'_1,\ldots,d'_k;m}$ are related by a sequence of castling transforms (of the form described above) and permutations of the dimensions. We say the datum $(d_1, \ldots, d_k; m)$ is *minimal* in its castling equivalence class if it minimises $\prod_{i=1}^{k} d_i$.

Lemma 3.12. *Consider the datum $(d_1, \ldots, d_k; m)$. Without loss of generality, we assume that $d_1 \leq d_2 \leq \cdots \leq d_k$. Let $N = m \cdot \prod_{i=1}^{k-1} d_i$. Then if $\frac{N}{2} < d_k < N$, the datum is not minimal in its castling equivalence class.*

Proof. We only need to show that if $\frac{N}{2} < d_k < N$, then the datum is not minimal. To see this, observe that we have a castling transform that takes $(d_1, \ldots, d_k; m)$ to $(d_1, \ldots, d_{k-1}, N - d_k; m)$, and the latter is smaller since $N - d_k < d_k$. □

Remark 3.13. If $d_1 = 1$, then $\rho_{d_1,d_2,\ldots,d_k;m}$ and $\rho_{d_2,\ldots,d_k;m}$ are equal up to isomorphism of the group and representation, so we can often assume without loss of generality that $d_i \geq 2$.

Even though it will not be relevant to us, we observe that each castling equivalence class contains a unique minimal datum (up to permutation). This follows from the fact that if any two data are related by (minimal) sequence of castling transforms, then the sequence of dimensions of representations produced by these transforms is monotonous, the proof of which is exactly the same as the proof of [28, Proposition 29].

4. Stability for tensor actions

In this section, we will prove Theorem 1.6, which gives a recursive characterisation of the generic stability properties for the tensor actions $\rho_{d_1,\cdots,d_k;m}$. Without loss of generality, we may assume that $d_1 \leq d_2 \leq \cdots \leq d_k$. By Corollary 3.10, we know that the properties we are looking at are invariant under the castling transform in part (3) of the theorem, so the majority of our work will be spent on the terminal cases. We now prove each part of the theorem separately.

For the first part, we need a simple lemma. It follows from the first fundamental theorem of invariant theory for the special linear group, a result that dates back to Weyl [43] but also has an elementary proof (see also [23, p. 7, Example]).

Lemma 4.1. *Consider the action of $G = \mathrm{SL}_d$ on $V = \mathrm{Mat}_{d,r}$ by left multiplication. If $d > r$, then every point $v \in V$ is G-unstable. In contrast, if $d \leq r$, then V is generically G-stable.*

Proof. Suppose $d > r$. Then any $v \in \mathrm{Mat}_{d,r}$ has rank at most r, so we can find $g \in \mathrm{SL}_d$ such that the range of gv is a subspace of the span of the first $r < d$ standard basis vectors. Then $\varphi(t) := g^{-1}\mathrm{diag}(t^{d-r}, \ldots, t^{d-r}, t^{-r}, \ldots, t^{-r})g \in \mathrm{SL}_d$ for all $t \neq 0$, and $\varphi(t)v \to 0$ as $t \to 0$.

Now suppose that $d \leq r$. By Lemma 2.9, it suffices to prove the claim in the case that $d = r$. Suppose $v \in \mathrm{Mat}_{d,d}$ is invertible (a Zariski-open set). Then its SL_d-orbit is equal to $\det^{-1}(\det v)$, hence closed. Since moreover its stabiliser is trivial, we conclude that V is generically stable.

Note that Lemma 2.9 was stated only for $\mathbb{F} = \mathbb{C}$. There are many ways to adapt the argument for $\mathbb{F} = \mathbb{R}$: for example, one can use [12, Proposition 2.23]. □

Proof of Theorem 1.6, part (1). As a representation of SL_{d_k}, the tensor space $\mathbb{F}^{d_1,\ldots,d_k;m}$ is isomorphic to $\mathrm{Mat}_{d_k,md_1d_2\cdots d_{k-1}}$, and hence every point is unstable by Lemma 4.1 since $d_k > d_1\cdots d_{k-1}m$. Hence every point is also unstable for the action of the larger group $G = \prod_{i=1}^{k}\mathrm{SL}_{d_i}$. □

For the second part, we will need the following result.

Lemma 4.2. *Let* $\pi\colon H \to \mathrm{SL}_d \subseteq \mathrm{GL}_d$ *be a d-dimensional representation of an algebraic group H. Consider the action of $G = H\times\mathrm{SL}_d$ on $\mathrm{Mat}_{d,d}$ given by $(h,g)\cdot A = \pi(h)Ag^{-1}$. For any full-rank matrix $A \in \mathrm{Mat}_{d,d}$, the stabiliser is given by $G_A = \{(h, A^{-1}\pi(h)A) \mid h \in H\}$. In particular, the stabiliser in general position is isomorphic to H.*

Proof. Straightforward. □

One point to note is that the kernel of the tensor action $\rho = \rho_{d_1,\ldots,d_k;m}$ is finite. So stability is equivalent to having a closed orbit and finite stabiliser. In particular, for $\mathbb{F} = \mathbb{C}$, Corollary 2.11 shows that generic stability of ρ is the same as the stabiliser in the general position being finite.

Proof of Theorem 1.6, part (2), for $\mathbb{F} = \mathbb{C}$. Let us define $H = \mathrm{SL}_{d_1} \times \mathrm{SL}_{d_2} \times \cdots \times \mathrm{SL}_{d_{k-1}}$ and $W = \mathbb{C}^{d_1,\ldots,d_{k-1};m}$. Then we can view $G \cong H \times \mathrm{SL}_{d_k}$ and $\mathbb{C}^{d_1,\ldots,d_k;m} \cong W \otimes \mathbb{C}^{d_k} \cong \mathrm{Mat}_{d_k,d_k}$, since $d_k = d_1\cdots d_{k-1}m$. So, the stabiliser in general position is H by Lemma 4.2, which is reductive. Hence, $\rho = \rho_{d_1,\ldots,d_k;m}$ is generically polystable by Theorem 2.10. As discussed above, the kernel of ρ is a finite group, so ρ is generically stable if and only if the stabiliser in general position H is finite. This happens precisely when $d_1 = d_2 = \cdots = d_{k-1} = 1$. □

Proof of Theorem 1.6, part (2), for $\mathbb{F} = \mathbb{R}$. This follows from [12, Proposition 2.23]. □

We already proved the third part of the theorem when we discussed the castling transforms of tensor actions.

Proof of Theorem 1.6, part (3). This follows from Corollary 3.10. □

We now prove the fourth and last part of the theorem, which is perhaps the most complicated. Here we wish to apply Theorems 2.12 and 2.13. Recall from Section 2.4 that the simple normal subgroups of $G = \mathrm{SL}_{d_1} \times \mathrm{SL}_{d_2} \times \cdots \times \mathrm{SL}_{d_k}$ are just $\mathrm{SL}_{d_1}, \mathrm{SL}_{d_2}, \ldots, \mathrm{SL}_{d_k}$. To compute the index of $V = \mathbb{C}^{d_1,\ldots,d_k;m}$ with respect to some SL_{d_i}, note that $V \cong (\mathbb{C}^{d_i})^{\oplus M}$ as an SL_{d_i}-representation, where $M = \frac{md_1\cdots d_k}{d_i}$. Now, the index of $\mathbb{C}^{d_i}$ with respect to SL_{d_i} is $\frac{1}{2d_i}$ and is additive. It follows that the index of $\mathbb{C}^{d_1,\ldots,d_k;m}$ with respect to SL_{d_i} is given by $\frac{M}{2d_i} = \frac{md_1d_2\cdots d_k}{2d_i^2}$. Since by assumption $d_1 \le d_2 \le \cdots \le d_k$, the smallest of these indices is the one for SL_{d_k}, given by $\frac{md_1d_2\cdots d_{k-1}}{2d_k}$. When $d_k \le \frac{1}{2}md_1d_2\cdots d_{k-1}$, as we assume in part (4) of the theorem, all indices therefore are at least one, so Theorems 2.12 and 2.13 are applicable. When $m = 1$, then the representation of G on V is irreducible. Elashvili has classified all irreducible representations of semisimple groups that are generically polystable but not generically stable. From the classification, one can extract the following (see [15, Theorem 9] and also [4, p. 9]).

Theorem 4.3 [15]. *Consider the irreducible representation $V = \mathbb{C}^{d_1,\ldots,d_k;1}$ of $G = \mathrm{SL}_{d_1}(\mathbb{C}) \times \cdots \times \mathrm{SL}_{d_k}(\mathbb{C})$. Assume that $2 \le d_1 \le \cdots \le d_k \le \frac{d_1\cdots d_{k-1}}{2}$. Then V satisfies the hypotheses of Theorem 2.13, hence is generically G-polystable. Moreover, V is not generically G-stable if and only if $k = 3$ and $(d_1, d_2, d_3) = (2, d, d)$ for some $d \ge 2$.*

Note that this result proves part (4) of the theorem when $\mathbb{F} = \mathbb{C}$ and $m = 1$. To deal with the case that $m \ge 2$, we will still make use of this theorem, together with a knowledge of the s.g.p.s.

For $(d_1, d_2, d_3) = (2, 2, 2)$, the stabiliser of $v = e_1^{\otimes 3} + e_2^{\otimes 3}$ is a s.g.p. It includes and has the same Lie algebra as the two-dimensional torus $\{(s, t, u) \in G : s, t, u \text{ diagonal}, stu = 1\}$.

For $(d_1, d_2, d_3) = (2, d, d)$, $d > 2$, the stabiliser of $v = e_1 \otimes I + e_2 \otimes A$, where I denotes the $d \times d$ identity matrix and A is a generic $d \times d$ diagonal matrix, is a s.g.p. It includes and has the same Lie algebra as the $(d-1)$-dimensional torus $\{(1, t, t^{-1}) \in G : t \text{ diagonal}\}$.

Proof of Theorem 1.6, part (4) for $\mathbb{F} = \mathbb{C}$. Since $d_k \leq \frac{d_1 \cdots d_{k-1} m}{2}$, the index of $V = \mathbb{C}^{d_1, \ldots, d_k; m}$ with respect to any simple normal subgroup of G is greater than or equal to one (as discussed above). When the inequality is strict, then ρ is generically stable by Theorem 2.12. Now suppose that $d_k = \frac{d_1 \cdots d_{k-1} m}{2}$. Then ρ is still generically polystable by Theorem 2.13. We now characterise when the representation is generically stable. If $d_1 = \cdots = d_{k-1} = 1$, then $d_k \leq \frac{m}{2} < m$, so ρ is generically stable by Lemma 4.1. Now assume that $d_{k-1} \geq 2$. Then $d_k = \frac{m d_1 \cdots d_{k-1}}{2} \geq m$. This means that if we consider the action of the larger group $H = \mathrm{SL}_{d_1} \times \cdots \times \mathrm{SL}_{d_k} \times \mathrm{SL}_m$ on $V = \mathbb{C}^{d_1} \otimes \cdots \otimes \mathbb{C}^{d_k} \otimes \mathbb{C}^m$, then the dimension d_k is still the largest among the dimensions $d_1, d_2, \ldots, d_k, m$. Accordingly, we can apply Theorem 4.3 to find that V is generically H-stable (hence also generically G-stable[4]), except if $(d_1, \ldots, d_k; m)$ is one of the following cases:

(a) $(1, \ldots, 1, 2, d, d; 1)$ for some $d \geq 2$,
(b) $(1, \ldots, 1, 2, 2; 2)$,
(c) $(1, \ldots, 1, d, d; 2)$ for some $d > 2$,
(d) $(1, \ldots, 1, 2, d; d)$ for some $d > 2$.

In case (a), we have $m = 1$ and hence $G \cong H$, so V is not generically G-stable either. To deal with the case that $m = 2$, we observe that an s.g.p. for G can be obtained by intersecting a generic H-conjugate of an s.g.p. for H with the subgroup G. From the description of the s.g.p.s above, we can observe the following. In case (b), the s.g.p. for G has dimension one (the dimension drops by one compared to H), while in case (c) it has dimension $d - 1$ (same as for the H-action). Thus we see that in either case, V is not generically G-stable. In contrast, in case (d), the s.g.p. for G is finite, so V is generically G-stable. □

Proof of Theorem 1.6, part (4) for $\mathbb{F} = \mathbb{R}$. This follows from [12, Proposition 2.23]. □

Finally, we need to prove that if ρ is not generically semistable, then it is unstable. For $\mathbb{F} = \mathbb{C}$, this statement is contained in Lemma 2.7. For $\mathbb{F} = \mathbb{R}$, it then follows from [12, Corollary 2.22 and Proposition 2.23]. This concludes the proof of Theorem 1.6.

5. A uniform characterization

In this section we prove Theorem 1.5, which gives a nonrecursive characterisation. Following [4], we define the following quantities for positive integers $k, d_1, \ldots, d_k$, and m:

$$R(d_1, \ldots, d_k; m) := m \prod_{i=1}^{k} d_i + \sum_{n=1}^{k} (-1)^n G_n(d_1, \ldots, d_k),$$

where

$$G_n(d_1, \ldots, d_k) := \sum_{1 \leq i_1 < \ldots < i_n \leq k} \gcd(d_{i_1}, \ldots, d_{i_n})^2,$$

as well as

$$g_{\max}(d_1, \ldots, d_k) := \max_{i < j} \gcd(d_i, d_j).$$

and

$$\Delta(d_1, \ldots, d_k; m) := m \prod_{i=1}^{k} d_i - 1 - \sum_{i=1}^{k} (d_i^2 - 1).$$

By convention, we define $g_{\max}(d) = 1$ for any $d \in \mathbb{Z}_{>0}$, and we always assume that $k \geq 1$.

---

[4]One way to see this is by using Corollary 2.11.

We saw earlier that generic semistability (polystability, stability) for tensor actions is symmetric in the d_is and invariant under the castling transform in part (3) of Theorem 1.6. It is also invariant under removing dimensions d_i equal to one.

It is not hard to verify that the quantities $R(d_1, \ldots, d_k; m)$, $g_{\max}(d_1, \ldots, d_k)$ and $\Delta(d_1, \ldots, d_k; m)$ have the same invariance properties. Hence, to prove Theorem 1.5, it suffices to consider the case when $(d_1, \ldots, d_k; m)$ is a minimal datum, and we may also assume that the d_i are sorted. Our analysis follows the same lines as the proof of [4, Proposition 5.3].

Lemma 5.1. *Suppose* $(d_1, \ldots, d_k; m)$ *is a minimal datum, and* $d_1 \leq d_2 \leq \cdots \leq d_k$. *Then*

1. $R < 0$ *if and only if* $d_k > m d_1 d_2 \cdots d_{k-1}$;
2. $R = 0$ *if and only if* $d_k = m d_1 d_2 \cdots d_{k-1}$;
3. $R > 0$ *if and only if* $d_k \leq \frac{1}{2} m d_1 d_2 \cdots d_{k-1}$.

Proof. According to Lemma 3.12, any minimal datum satisfies $d_k > m d_1 d_2 \cdots d_{k-1}$, $d_k = m d_1 d_2 \cdots d_{k-1}$ or $d_k \leq \frac{1}{2} m d_1 d_2 \cdots d_{k-1}$. If $d_1 = \cdots = d_k = 1$, then the lemma is immediate, since $R = m - 1$. Otherwise, we may assume that $d_1 \geq 2$ by removing all dimensions equal to one. We may also assume that $m \geq 2$, since when $m = 1$ the lemma is already proved in [4, Proposition 5.3]. Finally, observe that if we prove the 'if' directions for all three statements, then the 'only if' directions are automatic. Hence, we proceed to prove the 'if' directions in all three cases under the assumptions that $d_1 \geq 2$ and $m \geq 2$.

Let us write B_n for the terms in G_n that involve d_k and $A_n = G_n(d_1, \ldots, d_{k-1})$ for all other terms. Note that $A_k = 0$ and $B_1 = d_k^2$. Thus

$$R(d_1, \ldots, d_k; m) = m \prod_{i=1}^{k} d_i - d_k^2 + \sum_{n=1}^{k-1} (-1)^n (A_n - B_{n+1}). \tag{5.1}$$

Case (1): Suppose $d_k > d_1 \cdots d_{k-1} m$. Then $d_k = d_1 \cdots d_{k-1} m + \alpha$ for some $\alpha \geq 1$, and using equation (5.1),

$$R = -\alpha d_k + \sum_{n=1}^{k-1} (-1)^n (A_n - B_{n+1})$$

$$= -\alpha^2 - \alpha d_1 \cdots d_{k-1} m + \sum_{n=1}^{k-1} (-1)^n (A_n - B_{n+1}).$$

Clearly, $A_n \geq B_{n+1}$ for all n, so we can leave out the terms for odd n and obtain the bound

$$R \leq -\alpha^2 - \alpha d_1 \cdots d_{k-1} m + \sum_{n \geq 2 \text{ even}} (A_n - B_{n+1})$$

$$\leq -\alpha^2 - \alpha d_1 \cdots d_{k-1} m + \sum_{n \geq 2 \text{ even}} A_n$$

$$< -2 d_1 \cdots d_{k-1} + \sum_{n \geq 2 \text{ even}} A_n,$$

using that $m \geq 2$ and $\alpha \geq 1$. Now we are in the same situation as in [4, Eq. (9)] and find that $R < 0$.

Case (2): Suppose $d_k = d_1 \cdots d_{k-1} m$. Here we have $B_{n+1} = A_n$ for all n, so using equation (5.1),

$$R = m \prod_{i=1}^{k} d_i - d_k^2 = 0.$$

Case (3): Suppose $d_k \leq \frac{1}{2} d_1 \cdots d_{k-1} m$. If $k = 1$, then $d_1 \leq \frac{m}{2}$ and

$$R = md_1 - d_1^2 = d_1(m - d_1) \geq \frac{md_1}{2} > 0.$$

We now discuss the case that $k \geq 2$. Here,

$$
\begin{aligned}
R &= m \prod_{i=1}^{k} d_i - d_k^2 + \sum_{n=1}^{k-1} (-1)^n (A_n - B_{n+1}) \\
&= \frac{1}{4} d_1^2 \cdots d_{k-1}^2 m^2 - \left(\frac{1}{2} d_1 \cdots d_{k-1} m - d_k \right)^2 + \sum_{n=1}^{k-1} (-1)^n (A_n - B_{n+1}) \\
&\geq \frac{1}{4} d_1^2 \cdots d_{k-1}^2 m^2 - \left(\frac{1}{2} d_1 \cdots d_{k-1} m - d_{k-1} \right)^2 + \sum_{n=1}^{k-1} (-1)^n (A_n - B_{n+1}) \\
&= d_{k-1}^2 (d_1 \cdots d_{k-2} m - 1) + \sum_{n=1}^{k-1} (-1)^n (A_n - B_{n+1}),
\end{aligned}
$$

where the inequality follows because $d_{k-1} \leq d_k \leq \frac{1}{2} d_1 \cdots d_{k-1} m$. Leaving out the even terms, which are nonnegative since $A_n \geq B_{n+1}$, we obtain

$$
\begin{aligned}
R &\geq d_{k-1}^2 (d_1 \cdots d_{k-2} m - 1) + \sum_{n=1}^{k-1} (-1)^n (A_n - B_{n+1}) \\
&\geq d_{k-1}^2 (d_1 \cdots d_{k-2} m - 1) - \sum_{n \geq 1 \text{ odd}} (A_n - B_{n+1}) \\
&> d_{k-1}^2 (d_1 \cdots d_{k-2} m - 1) - \sum_{n \geq 1 \text{ odd}} A_n.
\end{aligned}
$$

Each of the $\binom{k-1}{n}$ GCDs contributing to A_n are $\leq d_{k-1}$, so

$$\sum_{n \geq 1 \text{ odd}} A_n \leq d_{k-1}^2 \sum_{n \geq 1 \text{ odd}} \binom{k-1}{n} = d_{k-1}^2 2^{k-2}$$

and hence

$$R > d_{k-1}^2 \left(d_1 \cdots d_{k-2} m - 1 - 2^{k-2} \right) \geq d_{k-1}^2 \left(2^{k-2} m - 1 - 2^{k-2} \right). \tag{5.2}$$

For $m \geq 2$ and $k \geq 2$, it holds that

$$2^{k-2} m - 1 - 2^{k-2} \geq 2^{k-2} - 1 \geq 0, \tag{5.3}$$

and hence we conclude that $R > 0$. □

Remark 5.2. Write $\mathcal{Z}(d_1, \ldots, d_k) := \sum_n (-1)^{n+1} \sum_{i_1 < i_2 < \cdots < i_n} \gcd(d_{i_1}, d_{i_2}, \ldots, d_{i_n})$. Then for $d_1, \ldots, d_k \in \mathbb{Z}_{\geq 1}$, one can interpret $\mathcal{Z}(d_1, \ldots, d_k)$ as the cardinality of $\bigcup_{i=1}^{k} \left(\mathbb{Z}[\frac{1}{d_i}]/\mathbb{Z} \right)$ in $\mathbb{Q}/\mathbb{Z}$. In particular, $\mathcal{Z}(d_1, \ldots, d_k) \geq 0$. Further, observe that $R(d_1, \ldots, d_k; m) = m \prod_{i=1}^{k} d_i - \mathcal{Z}(d_1^2, \ldots, d_k^2)$.

An alternate and short proof of the 'if'' statements in cases (1) and (2) in the above theorem is as follows. Observe that the quantity R is invariant under the transformation $(d_1, \ldots, d_k; m) \rightarrow (d_1, \ldots, d_{k-1}, d_k^*; m)$, where $d_k^* = m \prod_{i=1}^{k-1} d_i - d_k$ even in the case when some of the entries are negative or zero. Thus in case (1) we get $R(d_1, \ldots, d_k; m) = R(d_1, \ldots, d_{k-1}, d_k^*; m) < 0$ since

$md_1 \ldots d_{k-1}d_k^* < 0$ and $\mathcal{Z}(d_1^2, \ldots, d_{k-1}^2, (d_k^*)^2) \geq 0$. In case (2), using that $\mathcal{Z}(d_1^2, \ldots, d_{k-1}^2, 0) = 0$, one can deduce $R(d_1, \ldots, d_k; m) = R(d_1, \ldots, d_{k-1}, 0; m) = 0$.

Now we can prove Theorem 1.5.

Proof of Theorem 1.5. Generic semistability (polystability, stability) for tensor actions is invariant under the castling transform in part (3) of Theorem 1.6 and under permuting the dimensions d_i. The same is true for the quantities R, Δ, and $g_{\max}$. So, we can assume that $d_1 \leq d_2 \leq \cdots \leq d_k$ and that $(d_1, \ldots, d_k; m)$ is a minimal datum.

Case (1): Suppose $R > 0$. Then we know from Lemma 5.1 that $d_k \leq \frac{1}{2}md_1d_2 \cdots d_{k-1}$. If $k = 1$, then $d_1 \leq \frac{m}{2}$, so we must have $m \geq 2$. Further, $g_{\max} = 1$, so $R > 0$ implies that $R \geq g_{\max}^2$. Finally, ρ is always generically stable because the action of SL_{d_1} on $(\mathbb{C}^{d_1})^{\oplus m}$ is generically stable as long as $m \geq d_1$ (we have $m \geq 2d_1$). This concludes the proof in the case that $k = 1$.

Now we deal with $k \geq 2$. We may assume that $d_1 \geq 2$ by removing all dimensions equal to one (if all $d_i = 1$, then we can reduce to the case $k = 1$ discussed above). We distinguish two cases:

○ $m \geq 2$: In this case, we show that $R \geq g_{\max}^2$ and characterise equality. If $k > 2$, then equation (5.3) is not tight, and we see from equation (5.2) that

$$R > d_{k-1}^2 \geq g_{\max}^2.$$

For $k = 2$, we are in the matrix case. Since $2 \leq d_1 \leq d_2 \leq \frac{1}{2}md_1$, we find that

$$R(d_1, d_2; m) = md_1d_2 - d_1^2 - d_2^2 + \gcd(d_1, d_2)^2 = (md_1 - d_2)d_2 - d_1^2 + g_{\max}^2$$

$$\geq \frac{1}{2}md_1^2 - d_1^2 + g_{\max}^2 = \left(\frac{m}{2} - 1\right)d_1^2 + g_{\max}^2 \geq g_{\max}^2,$$

with equality if and only if $d_1 = d_2$ and $m = 2$, in which case also $g_{\max} = d_1 = d_2 \geq 2$. Thus we have proved that $R \geq g_{\max}^2$, with equality if and only if $k = 2, m = 2$ and $(d_1, d_2) = (d, d)$ for some $d \geq 2$. By part (4) of Theorem 1.6, this is precisely the case where ρ is generically polystable but not generically stable (when $d_k \leq \frac{1}{2}md_1d_2 \cdots d_{k-1}$ and $m \geq 2$).

○ $m = 1$: [4, Proposition 6.1] shows that in this case $\Delta \geq -2$, with equality precisely in the case that $k = 3$ and $(d_1, d_2, d_3) = (2, d, d)$ for some $d \geq 2$. (If $\Delta > -2$, then in fact $\Delta \geq 2$, but we do not need this.) By part (4) of Theorem 1.6, this is precisely the case where ρ is generically polystable but not generically stable (when $d_k \leq \frac{1}{2}md_1d_2 \cdots d_{k-1}$ and $m = 1$).

Case (2): Suppose $R = 0$. Then we know from Lemma 5.1 that $d_k = md_1d_2 \cdots d_{k-1}$. By part (2) of Theorem 1.6, ρ is generically polystable, and it is generically stable if and only if $d_1 = \cdots = d_{k-1} = 1$. When $k = 1$, we have $g_{\max} = 1$ (by definition), and this condition is always satisfied. Otherwise, $d_k = md_1d_2 \cdots d_{k-1}$ means $g_{\max} = \max_{i<j} \gcd(d_i, d_j) = \max_{i<k} d_i$. Thus, we find that in either case, $g_{\max} = 1$ if and only if ρ is generically stable.

Case (3): Suppose $R < 0$. By Lemma 5.1, we know that $d_k > md_1d_2 \cdots d_{k-1}$. Hence ρ is unstable by part (1) of Theorem 1.6. □

6. Maximum likelihood estimation for tensor normal models

In this section, we will prove Theorem 1.1, which characterises the boundedness of the likelihood function and the existence and uniqueness of MLEs for the tensor normal models.

The tensor normal models are the Gaussian group models corresponding to the tensor action. Thus the results on generic stability for tensor actions translate directly to results on maximum likelihood estimation for tensor normal models via Theorem 2.3. This connection is perfect for $\mathbb{F} = \mathbb{C}$, whereas more effort is required for $\mathbb{F} = \mathbb{R}$.

A technical point to note is that $G = \prod_{i=1}^k \mathrm{SL}_{d_i}$ is not a subset of $\mathrm{GL}(V)$, $V = \mathbb{F}^{d_1, \ldots, d_k; m}$, which is needed to apply Theorem 2.3 verbatim. However, this is a small issue, as we may simply replace G by

its homomorphic image $\rho_{d_1,\ldots,d_k;m}(G)$ and note that notions of semistability, polystability and stability are the same for both groups.

Proof of Theorem 1.1. We first consider the case of $\mathbb{F} = \mathbb{C}$. Consider the action of $G = \prod_{i=1}^{k} \mathrm{SL}_{d_i}(\mathbb{C})$ on $\mathbb{C}^{d_1,\ldots,d_k}$. The associated Gaussian group model is $\mathcal{M}_{\mathbb{C}}(d_1,\ldots,d_k)$. Thus, Corollary 2.8 implies that Theorem 1.5 translates precisely to Theorem 1.1.

We now discuss the relation between the real and the complex case. For both $\mathbb{F} = \mathbb{R}$ and $\mathbb{C}$, Theorem 1.5 shows that generic semistability is equivalent to generic polystability. Further, generic semistability (respectively, polystability) over $\mathbb{C}$ is equivalent to generic semistability (respectively, polystability) over $\mathbb{R}$; see [12, Proposition 2.23]. Finally, for both $\mathbb{F} = \mathbb{R}$ and $\mathbb{C}$, generic semistability is equivalent to almost sure boundedness of log-likelihood function because the semistable locus (over $\mathbb{F}$) is either empty or a (nonempty) Zariski-open subset (in particular, the complement of a measure zero subset); see [12, Corollary 2.15, Proposition 2.21, Corollary 2.22]. In fact, we claim that the following are equivalent:

1. $\rho_{d_1,\ldots,d_k;m}$ is generically semistable for $\mathbb{F} = \mathbb{C}$.
2. $\rho_{d_1,\ldots,d_k;m}$ is generically semistable for $\mathbb{F} = \mathbb{R}$.
3. $\rho_{d_1,\ldots,d_k;m}$ is generically polystable for $\mathbb{F} = \mathbb{C}$.
4. $\rho_{d_1,\ldots,d_k;m}$ is generically polystable for $\mathbb{F} = \mathbb{R}$.
5. For the tensor normal model $\mathcal{M}_{\mathbb{C}}(d_1,\ldots,d_k)$, we have almost sure boundedness of log-likelihood function for m samples.
6. For the tensor normal model $\mathcal{M}_{\mathbb{R}}(d_1,\ldots,d_k)$, we have almost sure boundedness of log-likelihood function for m samples.
7. For the tensor normal model $\mathcal{M}_{\mathbb{C}}(d_1,\ldots,d_k)$, an MLE exists almost surely for m samples.
8. For the tensor normal model $\mathcal{M}_{\mathbb{R}}(d_1,\ldots,d_k)$, an MLE exists almost surely for m samples.

The equivalence of (1)–(6) was discussed above. The implications (3) $\implies$ (7) and (4) $\implies$ (8) follow from Theorem 2.3 since the complement of a Zariski-open subset has Lebesgue measure zero. Further, it is also immediate that (7) $\implies$ (5) and (8) $\implies$ (6). This shows the equivalence of all eight statements.

Moreover, $\rho_{d_1,\ldots,d_k;m}$ is generically stable over $\mathbb{F} = \mathbb{C}$ if and only if the same holds for $\mathbb{F} = \mathbb{R}$; see again [12, Proposition 2.23]. In either case, generic stability implies the almost sure existence of a unique MLE by Theorem 2.3. However, the converse is not necessarily true when $\mathbb{F} = \mathbb{R}$, and this is what needs to be investigated.

To summarise, the only cases we need to study further are those in which $\rho_{d_1,\ldots,d_k;m}$ is generically polystable but not generically stable. According to Theorem 1.6, these are the castling equivalence classes of the minimal data below:

1. $d_k = m d_1 d_2 \cdots d_{k-1}$ and $d_1 \cdots d_{k-1} > 1$.
2. $(d_1,\ldots,d_k,m) = (1,1,\ldots,1,d,d;2)$ with $d \geq 2$.
3. $(d_1,\ldots,d_k,m) = (1,1,\ldots,1,2,d,d;1)$ with $d \geq 2$.

To conclude the proof of Theorem 1.1, we need to show for these that we do not have the almost sure existence of a unique MLE also over $\mathbb{F} = \mathbb{R}$. By Corollary 2.5 and Corollary 3.9, it suffices to prove that in any of these three minimal cases, there is a Euclidean open subset consisting of points with noncompact stabilisers for each of the above minimal data. Note that Euclidean open subsets have positive Lebesgue measure.

For case (1), observe that the proof of Lemma 4.2 works even when the underlying field is $\mathbb{R}$. So, in fact, there is a nonempty Zariski-open subset of V (in particular, a set of positive measure) where the stabiliser is isomorphic to $\prod_{i=1}^{k-1} \mathrm{SL}_{d_i}(\mathbb{R})$, which is noncompact unless $d_1 = \cdots = d_{k-1} = 1$.

We now address case (2) and distinguish two cases:

○ $d \geq 3$: For generic $v \in \mathrm{Mat}_{d,d}^2 = (\mathbb{R}^d \otimes \mathbb{R}^d)^{\oplus 2}$, we give a sequence of elements in the stabiliser with no convergent subsequence (hence proving that the stabiliser is not compact). It was proved in [12,

Lemma 6.2] that for generic $v \in \mathrm{Mat}_{d,d}^2$, there exists $(g, h) \in G_v$ such that g and h have eigenvalues with absolute value not equal to 1. Since $G_v \subseteq \mathrm{SL}_d \times \mathrm{SL}_d$, this means $\{(g^n, h^n)\}_{n \in \mathbb{Z}_{>0}}$ is a sequence of elements in G_v with no convergent subsequence. Hence G_v is not compact. This gives a Zariski open subset consisting of points with noncompact stabiliser.

○ $d = 2$: It is easy to see that the stabiliser of $v_{a,b} = \left(\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right), \left(\begin{smallmatrix} a & 0 \\ 0 & b \end{smallmatrix}\right)\right) \in \mathrm{Mat}_{2,2}^2$ is not compact for any $a, b \in \mathbb{R}$ (compare the discussion below Theorem 4.3). Now, let us consider $W = \{(A, B) \in \mathrm{Mat}_{2,2}^2 \mid \det(A) \neq 0, \det(tI - A^{-1}B) \text{ has distinct real roots}\}$. Then it is easy to see that every $w \in W$ is in the $\mathrm{SL}_d \times \mathrm{SL}_d$ orbit of $v_{a,b}$ for an appropriate choice of a and b (indeed, just the eigenvalues of $A^{-1}B$). Next, observe that W is a full-dimensional semialgebraic set; indeed, it is described by one Zariski-open conditions ($\det(A) \neq 0$) and one inequality (the discriminant of $\det(tI - A^{-1}B)$ is larger than zero). Thus, W is an Euclidean-open subset (hence, a set of positive Lebesgue measure), and every point in W has a noncompact stabiliser.

Finally, case (3) follows from case (2) in view of Lemma 6.1 below. □

Lemma 6.1. *Let $H \subseteq G$ be a closed subgroup of an algebraic group, and let V be a rational representation of G. Let $v \in V$. If G_v is compact, then so is H_v.*

Proof. $H_v = G_v \cap H$ is a closed subset of G_v and hence compact if G_v is compact. □

We end this section with a proof of Corollary 1.3.

Proof of Corollary 1.3. Since Theorem 1.1 does not differentiate between $\mathbb{F} = \mathbb{R}$ and $\mathbb{F} = \mathbb{C}$, it suffices to prove this in the case of $\mathbb{F} = \mathbb{C}$. Here, statistical notions correspond precisely to stability notions by Corollary 2.8, so we will make our arguments in the language of stability. First, observe that $\lceil r \rceil \leq \mathrm{mlt}_b (= \mathrm{mlt}_e)$ because $\rho_{d_1,\ldots,d_k;m}$ is unstable unless $m \geq r$ by part (1) of Theorem 1.6.

Now, let $c = \lceil r \rceil$, so $d_k = cd_1 \cdots d_{k-1} - \alpha$ for some $0 \leq \alpha < d_1 d_2 \cdots d_{k-1}$. To show $\mathrm{mlt}_u \leq c + 1$, it suffices to show that $\rho_{d_1,\ldots,d_k,c+1}$ is generically stable by Lemma 2.9.

We see that $\rho_{d_1,\ldots,d_k;c+1}$ is castling equivalent to $\rho_{d_1,\ldots,d_{k-1},d_1 d_2 \cdots d_{k-1}+\alpha;c+1}$. It suffices to show that one of them is generically stable. Observe that both $A = d_k$ and $B = d_1 d_2 \cdots d_{k-1} + \alpha$ are larger than d_{k-1}, so the dimensions are already in order. Since $A + B = (c + 1)d_1 \cdots d_{k-1}$, we get that either A or B is $\leq \frac{1}{2}(c+1)d_1 \cdots d_{k-1}$. Hence, we get generic stability for $\rho_{d_1,\ldots,d_k;c+1}$ by parts (3) and (4) of Theorem 1.6 unless $(d_1, \ldots, d_k; c + 1)$ (or $(d_1, \ldots, d_{k-1}, B; c + 1)$) is one of $(2, d, d; 1)$ or $(d, d; 2)$. The former is not possible because $c + 1 \geq 2$, and the latter is not possible because $k \geq 3$ by assumption. □

7. Dimension of the GIT quotient

In this section, let the underlying field be $\mathbb{F} = \mathbb{C}$. Let V be a rational representation of a reductive group G. Then the GIT quotient $\mathbb{P}V /\!\!/ G$ is defined as $\mathrm{Proj}(\mathbb{C}[V]^G)$, the projective variety associated to the ring of invariants (with its natural grading).

Given what we have computed, we can also compute the dimension of the GIT quotient for the action of $G = \prod_i \mathrm{SL}_{d_i}$ on $V = \mathbb{F}^{d_1,\ldots,d_k;m}$. This relies on Rosenlicht's theorem [33, Theorem 2] (see also the proof of [4, Lemma 3.1]).

Theorem 7.1 (Rosenlicht). *Let V be a rational representation of a connected semisimple group G. Let H be the stabiliser in general position. Then $\dim(\mathbb{P}V /\!\!/ G) = \dim(\mathbb{P}V) - \dim(G) + \dim(H)$, where $\dim(\mathbb{P}V /\!\!/ G) = -1$ if and only if $\mathbb{P}V /\!\!/ G = \emptyset$.*

For the tensor action, this means

$$\dim(\mathbb{P}V /\!\!/ G) = \Delta(d_1, \ldots, d_k; m) + \dim H, \tag{7.1}$$

where $\Delta = \Delta(d_1, \ldots, d_k; m) = m \prod_{i=1}^k d_i - 1 - \sum_{i=1}^k (d_i^2 - 1)$ as defined above and where H is the stabiliser in general position.

Proof of Theorem 1.7. By Lemma 3.6, the dimension of the GIT quotient is invariant under castling transforms, so we may assume that $(d_1, \ldots, d_k; m)$ is minimal. We handle each case separately.

Case (1): Suppose $R < 0$. Then ρ is unstable by Theorem 1.5. This means the invariant ring is given by $\mathbb{C}[V]^G = \mathbb{C}$ and $\mathbb{P}V/\!\!/G$ is empty.

Case (2): Suppose $R = 0$. Then $md_1d_2 \cdots d_{k-1} = d_k$ by Lemma 5.1. We identify $V \cong \mathrm{Mat}_{d_k, d_k}$. For the left-right action of $\mathrm{SL}_{d_k} \times \mathrm{SL}_{d_k}$, the ring of invariants is $\mathbb{C}[\det]$, where det denotes the determinant polynomial. The same is true when we restrict to the second SL_{d_k}, say. Since $\{1\} \times \mathrm{SL}_{d_k} \subseteq G \subseteq \mathrm{SL}_{d_k} \times \mathrm{SL}_{d_k}$, the ring of invariants for ρ, is also $\mathbb{C}[\det]$. Thus, $\mathbb{P}V/\!\!/G$ is a single point.

Case (3): Suppose $R > 0$. Whenever ρ is generically stable, equation (7.1) implies that the dimension of the GIT quotient is Δ (recall that the kernel of ρ is zero-dimensional), while if ρ is only generically polystable, we need to add the dimension of the stabiliser in general position. There are two cases to consider:

- $m = 1$ and $\Delta = -2$: In this case, $k = 3$ and $(d_1, d_2, d_3) = (2, d, d)$ for some $d \geq 2$, as we saw in the proof of Theorem 1.5. If $d = 2$, then the s.g.p. is two-dimensional, while if $d > 2$, it is $(d-1)$-dimensional (see the proof of Theorem 1.6, part (4)). Thus, since $g_{\max} = d$,

$$\dim(\mathbb{P}V/\!\!/G) = \Delta + \dim H = -2 + \begin{cases} 2 & \text{if } g_{\max} = 2 \\ g_{\max} - 1 & \text{if } g_{\max} > 2 \end{cases} = \max(g_{\max} - 3, 0),$$

  which is also contained in [4, Theorem 1.2].

- $m = 2$ and $R = g_{\max}^2 > 1$: In this case, similarly, $k = 2$ and $(d_1, d_2) = (d, d)$ for some $d \geq 2$, again by the proof of Theorem 1.5. If $d = 2$, then the s.g.p. is one-dimensional, while if $d > 2$, then the s.g.p. is $(d-1)$-dimensional (see proof of Theorem 1.6, part (4)). Thus $\dim H = g_{\max} - 1$ in either case, and hence

$$\dim(\mathbb{P}V/\!\!/G) = \Delta + \dim H = 1 + (g_{\max} - 1) = g_{\max}. \qquad \square$$

References

[1] Z. Allen-Zhu, A. Garg, Y. Li, R. Oliveira and A. Wigderson, *Operator scaling via geodesically convex optimization, invariant theory and polynomial identity testing*, STOC'18—Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, 172–181, ACM, New York, 2018.

[2] C. Amendola, K. Kohn, P. Reichenbach and A. Seigal, *Invariant theory and scaling algorithms for maximum likelihood estimation*, SIAM J. Appl. Algebra Geometry, **5** (2021), 304–337.

[3] E. M. Andreev, É. B. Vinberg and A. G. Élashvili, *Orbits of greatest dimension in semi-simple linear Lie groups*, Funktsional. Anal. i Prilozhen **1** (1967), 3–7.

[4] J. Bryan, Z. Reichstein and M. Van Raamsdonk, *Existence of locally maximally entangled quantum states via geometric invariant theory*, Ann. Henri Poincaré **19** (2018), 2491–2511.

[5] J. Bryan, S. Leutheusser, Z. Reichstein and M. Van Raamsdonk, *Locally maximally entangled states of multipart quantum systems*, Quantum **3** (2019), 115.

[6] P. Bürgisser, A. Garg, R. Oliveira, M. Walter and A. Wigderson, *Alternating minimization, scaling algorithms, and the null-cone problem from invariant theory*, 9th Innovations in Theoretical Computer Science, no. 24, 20 pp., LIPIcs. Leibniz Int. Proc. Inform., **94**, Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2018.

[7] P. Bürgisser, C. Franks, A. Garg, R. Oliveira, M. Walter and A. Wigderson, *Towards a Theory of Non-Commutative Optimization: Geodesic 1st and 2nd Order Methods for Moment Maps and Polytopes*, 60th Annual IEEE Symposium on Foundations of Computer Science – FOCS 2019, 845–861, IEEE Computer Soc., Los Alamitos, CA, 2020.

[8] H. Derksen and V. Makam, *Polynomial degree bounds for matrix semi-invariants*, Adv. Math. **310** (2017), 44–63.

[9] H. Derksen and V. Makam, *Generating invariant rings of quivers in arbitrary characteristic*, J. Algebra **489** (2017), 435–445.

[10] H. Derksen and V. Makam, *Algorithms for orbit closure separation for invariants and semi-invariants of matrices*, to appear in Algebra and Number Theory.

[11] H. Derksen and V. Makam, *An exponential lower bound for the degrees of invariants of cubic forms and tensor actions*, Adv. Math. **368** (2020), 107136.

[12] H. Derksen and V. Makam, *Maximum likelihood estimation for matrix normal models via quiver representations*, SIAM J. Appl. Algebra Geometry, **5** (2021), 338–365.

[13] M. Drton, S. Kuriki and P. Hoff, *Existence and Uniqueness of the Kronecker Covariance MLE*, arXiv:2003.06024 [math.ST], 2020.

[14] P. Dutilleul, *The MLE algorithm for the matrix normal distribution*, J. Statist. Comput. Simul. **64** (1999), 105–123.

[15] A. G. Élashvili, *Stationary subalgebras of points of general position for irreducible linear Lie groups*, Funkcional. Anal. i Priložen **6** (1972), no. 2, 65–78.

[16] C. Franks, R. Oliveira, A. Ramachandran, M. Walter, *Near optimal sample complexity for matrix and tensor normal models via geodesic convexity*, arXiv:2110.07583

[17] W. Fulton, *Young Tableaux*, With applications to representation theory and geometry, London Mathematical Society Student Texts **35**, Cambridge University Press, Cambridge, 1997, x+260 pp.

[18] A. Garg, L. Gurvits, R. Oliveira and A. Widgerson, *A deterministic polynomial time algorithm for non-commutative rational identity testing*, 57th Annual IEEE Symposium on Foundations of Computer Science–FOCS 2016, 109–117, IEEE Computer Soc., Los Alamitos, CA, 2016.

[19] G. Ivanyos, Y. Qiao and K. V. Subrahmanyam, *Non-commutative Edmonds' problem and matrix semi-invariants*, Comput. Complexity **26** (2017), no. 3, 717–763.

[20] G. Ivanyos, Y. Qiao and K. V. Subrahmanyam, *Constructive non-commutative rank computation is in deterministic polynomial time*, Comput. Complexity **27** (2018), no. 4, 561–593.

[21] V. G. Kac, *Infinite root systems, representations of graphs and invariant theory*, Invent. Math. **56** (1980), no. 1, 57–92.

[22] A. Klyachko, *Dynamical symmetry approach to entanglement*, NATO Security through Science Series D **7**, 2007.

[23] H. Kraft, C. Procesi, *Classical Invariant Theory: A primer*. Lecture notes, July 1996.

[24] S. Koga and S. Y. Zhang, *Inter-tree and intra-tree variations in ring width and wood density components in Balsam fir (Abiesbalsamea)*, Wood Sci. Technol. **38** (2004) 149–162.

[25] N. Lu and D. Zimmerman, On likelihood-based inference for a separable covariance matrix, Technical Report 337, Statistics and Actuarial Science Dept., Univ. of Iowa, Iowa City, IA, 2004.

[26] N. Lu and D. Zimmerman, *The likelihood ratio test for a separable covariance matrix*, Statistics and Probability Letters **73** (2005), 449–457.

[27] A. M. Manceur and P. D. Dutilleul, *Maximum likelihood estimation for the tensor normal distribution: algorithm, minimum sample size, and empirical bias and dispersion*, J. Comput. Appl. Math. **239** (2013), 37–49.

[28] L. Manivel, *Prehomogeneous spaces and projective geometry*, Rend. Semin. Mat. Univ. Politec. Torino **71** (2013), no. 1, 35–118.

[29] A. M. Popov, Finite isotropy subgroups in general position of irreducible semisimple linear Lie groups (Russian) Trudy Moskov. Mat. Obshch. **50** (1987), 209–248, 262; translation in Trans. Moscow Math. Soc. 1988, 205–249.

[30] V. L. Popov, *Stability criteria for the action of a semisimple group on a factorial manifold*, Math. USSR-Izvestiya **4** (1970), 527–535.

[31] C. Procesi, Lie Groups, An approach through invariants and representations. Universitext. Springer, New York, 2007. xxiv+596 pp.

[32] B. Roś, B. Fetsje, J. C. de Munck and Mathisca C. M. de Gunst, *Existence and uniqueness of the maximum likelihood estimator for models with a Kronecker product covariance structure*, J. Multivariate Anal. **143** (2016), 345–361.

[33] M. Rosenlicht, *Some basic theorems on algebraic groups*, Amer. J. Math. **78** (1956), 401–443.

[34] A. Roy, R. Leiva, *Likelihood ratio tests for triply multivariate data with structured correlation on spatial repeated measurements*, Statist. Probab. Lett. **78** (2008) 1971–1980.

[35] M. Sato and T. Kimura, *A classification of irreducible prehomogeneous vector spaces and their relative invariants*, Nagoya Math. J. **65** (1977), 1–155.

[36] I. Soloveychik, D. Trushin, *Gaussian and robust Kronecker product covariance estimation: Existence and uniqueness*, Journal of Multivariate Analysis **149** (2016), 92–113.

[37] M. S. Srivastava, T. von Rosen, D. von Rosen, *Models with a Kronecker product covariance structure: estimation and testing*, Math. Methods Statist. **17** (2008), no. 4, 357–370.

[38] È. B. Vinberg and V. L. Popov, Invariant theory (Russian) Algebraic geometry, **4** (Russian), 137–314, 315, Itogi Nauki i Tekhniki, Sovrem. Probl. Mat. Fund. Naprav., 55, Akad. Nauk SSSR, Vsesoyuz. Inst. Nauchn. i Tekhn. Inform., Moscow, 1989.

[39] F. Venturelli, *Prehomogeneous tensor spaces*, Linear Multilinear Algebra **67** (2019), no. 3, 510–526.

[40] M. Walter, B. Doran, D. Gross, M. Christandl, *Entanglement polytopes: multiparticle entanglement from single-particle information* Science **340** (2013), no. 6137, 1205–1208.

[41] M. Walter, Multipartite quantum states and their marginals, PhD thesis, ETH Zurich (2014).

[42] K. Werner, M. Jansson, and P. Stoica, *On estimation of covariance matrices with Kronecker product structure*, IEEE Transactions on Signal Processing **56** (2008), 478–491.

[43] H. Weyl, The Classical Groups, their Invariants and Representations, Princeton Mathematical Series, vol. **1**, Princeton University Press, Princeton, 1946.

[44] J. Weyman, Cohomology of Vector Bundles and Syzygies, Cambridge Tracts in Mathematics, vol. **149**, Cambridge University Press, Cambridge, 2003, xiv+371 pp.