Minimality of tensors of fixed multilinear rank

Alexander Heaton\textsuperscript{a}, Khazhgali Kozhasov\textsuperscript{b} and Lorenzo Venturello\textsuperscript{c}

\textsuperscript{a}Max Planck Institute for Mathematics in the Sciences, Technische Universität Berlin, Leipzig; \textsuperscript{b}Technische Universität Braunschweig, Institut für Analysis und Algebra, Braunschweig, Germany; \textsuperscript{c}Max Planck Institute for Mathematics in the Sciences, Leipzig

Abstract

We discover a geometric property of the space of tensors of fixed multilinear (Tucker) rank. Namely, it is shown that real tensors of the fixed multilinear rank form a minimal submanifold of the Euclidean space of tensors endowed with the Frobenius inner product. We also establish the absence of local extrema for linear functionals restricted to the submanifold of rank-one tensors, finding application in statistics.

1. Introduction

In the following, by $\mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_d}$ we denote the space of real $(n_1, \ldots, n_d)$-tensors. Upon fixing a basis $E^j = \{e^j_1, \ldots, e^j_{n_j}\}$ of $\mathbb{R}^{n_j}$ for $j = 1, \ldots, d$, a tensor $T = \sum_{j=1}^d \sum_{i_j=1}^{n_j} t_{i_1 \ldots i_d} e^1_{i_1} \otimes \cdots \otimes e^d_{i_d}$ is identified with the array $T = (t_{i_1 \ldots i_d})$ of its coordinates with respect to the induced basis $\{e^1_{i_1} \otimes \cdots \otimes e^d_{i_d}\}$ of $\mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_d}$. The space $\mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_d}$ is endowed with the Frobenius inner product defined by

$$\langle T, S \rangle = \sum_{j=1}^d \sum_{i_j=1}^{n_j} t_{i_1 \ldots i_d} s_{i_1 \ldots i_d},$$

where $T = (t_{i_1 \ldots i_d})$, $S = (s_{i_1 \ldots i_d})$ are written with respect to (any) orthonormal bases $E^1, \ldots, E^d$ of Euclidean spaces $\mathbb{R}^{n_1}, \ldots, \mathbb{R}^{n_d}$. In the following, we are fixing such orthonormal bases and our computations are carried out with the respective coordinate expressions of tensors. It is possible to define the Frobenius product with respect to arbitrary (not necessarily orthonormal) bases (see e.g. [1, (4.64)]) of the underlying Euclidean spaces. However, in this case, the corresponding formula for the inner product is less practical than (1).
The multilinear rank of $T \in \mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_d}$ is the tuple

$$\mu \text{rank}(T) = \min \left\{ (r_1, \ldots, r_d) \in \mathbb{N}_0^d : T \in W_1 \otimes \cdots \otimes W_d, W_j \subseteq \mathbb{R}^{n_j}, \dim W_j = r_j \right\},$$

where $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$ and each $W_j$ is a linear subspace of $\mathbb{R}^{n_j}$. The number $r_j$ is equal to the rank of the $n_j \times \prod_{k \neq j} n_k$ matrix obtained by flattening or unfolding the tensor along mode $j$ [2, Theorem 6.13]. In the case $d = 2$ the multilinear rank $(r_1, r_2)$ of a matrix $T$ satisfies $r_1 = r_2 = \text{rank}(T)$, since the rank of the row and column space of any matrix coincide.

For tensors with more than 2 modes, the multilinear rank is different than the classical rank

$$\text{rank}(T) = \min \left\{ r \in \mathbb{N}_0 : T = \sum_{i=1}^r v_{1,i} \otimes \cdots \otimes v_{d,i}, \quad v_{j,i} \in \mathbb{R}^{n_j} \right\}.$$

The latter notion of rank is also important for many applications (see [3]), but will not be the focus of this article. For $d \geq 1$ and $r \in \mathbb{N}_0^d$, let $T_{d,r}$ be the set of order $d$ tensors with multilinear rank $r$. It is well known (see, for example, [4]) that this set is a smooth submanifold of $\mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_d}$ of dimension

$$\dim(T_{d,r}) = \sum_{j=1}^d r_j (n_j - r_j) + \prod_{j=1}^d r_j.$$

The topology of $T_{d,r}$ was recently studied in [5]. They show, for example, that $T_{d,r}$ is path-connected unless $n_j = r_j = \prod_{k \neq j} r_j$ for some $j$. The notion of multilinear rank was introduced in [3] and popularized in [6] where it was used to show that the Tucker decomposition (see [7]) provides a convincing generalization of the matrix singular value decomposition (SVD). The set $T_{d,r}$ and the Tucker decomposition in general have been used in numerous interesting applications, including the technique TensorFaces in computer vision [8]. Note that throughout this article we refer to the set of tensors of the fixed multilinear rank, rather than the related subspace variety consisting of all tensors with $\mu \text{rank}(T) \leq (r_1, \ldots, r_d)$, which is also a well-studied object of interest [9, Chapter 7].

A common task is to find the best low-rank approximation of a tensor [10]. For matrices, the celebrated Eckart-Young theorem [11] states that truncating the SVD yields a closed-form solution, but for tensor rank the problem is ill-posed [12] due to the phenomenon of border rank. The subspace variety is Zariski closed, providing a nice setting for existence, but is not smooth. As a consequence, computing the best multilinear approximation over the subspace variety becomes more difficult, since not all points admit tangent spaces.

In contrast, the problem of finding the best multilinear low-rank approximation is well posed on $T_{d,r}$ with a unique solution with respect to the Frobenius norm [12, Corollary 4.5]. The manifold structure of low-rank tensors has been used in numerical analysis and computational physics (see [13] for an overview) and, in general, methods of Riemannian optimization can be utilized [14,15]. In recent work [16], the authors studied the problem of tensor completion, filling in the missing entries of a tensor to achieve a low-rank tensor. Tensors of fixed multilinear rank $T_{d,r}$ were used due to their manifold structure, implementing a version of the nonlinear conjugate gradient method, see also [17]. $T_{d,r}$ was also used in [18,19] to formulate the tensor approximation problem over a product of Grassmann manifolds. In [20], the manifold structure of $T_{d,r}$ was used to derive nonlinear
differential equations whose solution is a time-varying family of tensors $S(t) \in T_{d,r}$ which provide the best multilinear rank approximation to a given family $T(t) \in \mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_d}$, a problem called dynamical tensor approximation. Therefore, the geometry of the set $T_{d,r}$ is important for applications.

Our main result discovers a new geometric property of $T_{d,r}$, namely its minimality.

**Theorem 1.1:** For any $r \in \mathbb{N}_0^d$ the manifold $T_{d,r}$ is a minimal submanifold of the Euclidean space $(\mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_d}, \langle \cdot, \cdot \rangle)$, that is, its mean curvature vector field is identically zero.

Minimal submanifolds are mathematical models of soap films: they minimize the volume locally around every point. When $d = 2$ the manifold $T_{2,r}$, $r = (r,r)$, consists of $n_1 \times n_2$ matrices of rank $r$. Its minimality was recently proved in [21,22] and for $n_1 = n_2 = r + 1$ earlier in [23]. Thus, Theorem 1.1 generalizes this property from matrices to higher order tensors.

The case $r = (1, \ldots, 1)$ is of particular interest, since the elements of $T_{d,(1,\ldots,1)}$ are precisely tensors of rank 1. The manifold $T_{d,(1,\ldots,1)}$ is the (affine) cone over the real part of the (projective) Segre variety. In what follows we use the term ‘Segre variety’ when referring to the manifold $T_{d,(1,\ldots,1)}$ and denote it for simplicity by $S_d$. If we slice $T_{d,(1,\ldots,1)}$ with the affine hyperplane $A$ containing the tensors with sum of coordinates equal to 1 and consider its nonnegative part, we obtain a statistical model $\mathcal{M}$ which parametrizes $d$-tuples of independent discrete random variables $(X_1, \ldots, X_d)$. The tensor $T = (t_{i_1 \ldots i_d})$ expresses the joint probability $t_{i_1 \ldots i_d} = \mathbb{P}(X_1 = i_1, \ldots, X_d = i_d)$, and the condition of having rank 1 corresponds to statistical independence of $X_1, \ldots, X_d$. The popular Wasserstein distance between two probability distributions arises from optimal transport, measuring how much work is required to ‘move’ one distribution to the other. In this discrete setting, the Wasserstein distance between a distribution $T \in A$ and $\mathcal{M}$ is the smallest scaling factor for which a certain polyhedron centred at $T$ intersects $\mathcal{M}$. The closest points on $\mathcal{M}$ to $T$ are those in the intersection. In [24, Section 6] it was experimentally observed that the optimal solution to this problem is never attained in the relative interior of a face of maximal dimension. One possible explanation for this experimental observation would be that the restriction of a linear functional to $\mathcal{M}$ attains no local minima (respectively maxima) in its relative interior. Corollary 1.4 below shows that it is indeed the case. This in turn follows from the following related result for the Segre variety $S_d = T_{d,(1,\ldots,1)}$. In the statement of this theorem, $\vec{\ell}$ is a vector in $\mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_d}$ representing a linear functional.

**Theorem 1.2:** Let $T \in S_d$ and let $P_\ell = \{ T + V : \langle \vec{\ell}, V \rangle = 0 \} \subset \mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_d}$ be an affine hyperplane containing the tangent space to $S_d$ at $T$. Then no neighbourhood of $T$ in $S_d$ is completely contained in one of the half spaces $P_\ell^\pm = \{ T + V : \pm \langle \vec{\ell}, V \rangle \geq 0 \}$. In particular, the linear functional $L_\ell(S) = \langle \vec{\ell}, S \rangle$ does not attain a local maximum or minimum on $S_d$.

A minimal surface in $\mathbb{R}^3$ with non-degenerate second fundamental form has saddle-like shape locally around every point. In particular, no linear functional can be minimized or maximized on such a surface. An analogous fact is true for minimal submanifolds in $\mathbb{R}^n$ satisfying some non-degeneracy conditions, see Remark 4.2. In view of this and the minimality of $T_{d,r}$ (see Theorem 1.1), the assertion of Theorem 1.2 is anticipated. However
to prove it, in Section 4 we have to deal with finer geometric information of $S_d$ than just its mean curvature.

**Corollary 1.3:** Let $A$ be an affine hyperplane in $\mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_d}$, and let $L: \mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_d} \to \mathbb{R}$ be a linear functional not constant on $A$. Then the restriction of $L$ to $S_d \cap A$ has no local minima (respectively maxima).

If we take $A$ to be the affine hyperplane of $\mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_d}$ consisting of tensors with entries summing up to 1, we obtain the following application in statistics.

**Corollary 1.4:** Let $\mathcal{M}$ be the statistical model of $d$ independent discrete random variables. Then any linear functional which is not constant on $\mathcal{M}$ has no local minimum (or maximum) on the relative interior of $\mathcal{M}$.

## 2. Preliminaries

This section reviews basic concepts we will need. For more on tensors and tensor decompositions, see [9], while for differential geometry see [25,26]. We first recall a definition of the mean curvature vector field of a submanifold of an Euclidean space, see [26, Ch. VII] for more details.

Let $(\mathbb{R}^n, \langle \cdot, \cdot \rangle)$ be an Euclidean space and let $M \subset \mathbb{R}^n$ be a smooth $m$-dimensional submanifold. Consider a local parametrization $\gamma: U \to M$, where $U$ is an open subset of $\mathbb{R}^m$. The first-order partial derivatives $\partial_{u_1} \gamma(u), \ldots, \partial_{u_m} \gamma(u)$ form a basis of the tangent space $T_{\gamma(u)} M$ of $M$ at $\gamma(u)$ and we denote by $G_{\gamma(u)} = \langle (\partial_{u_i} \gamma(u), \partial_{u_j} \gamma(u)) \rangle$ the Gram matrix of the metric of $M$ with respect to this basis. Thus, $G_{\gamma(u)}$, $u \in U$, is a smooth field of positive-definite matrices along $r(U) \subset M$. The **second fundamental form of** $M$ is a symmetric bilinear form $b$ on the tangent bundle $TM$ to $M$ with values in the normal bundle $(TM)^\perp \subset \mathbb{R}^n$ to $M$ defined at a point $r(u) \in M$ via

$$ b(u, v) = \sum_{i,j=1}^{m} u^i v^j \langle \partial_{u_i u_j}^2 \gamma(u) \rangle, $$

where $u = \sum_{i=1}^{m} u^i \partial_{u_i} \gamma(u)$, $v = \sum_{j=1}^{m} v^j \partial_{u_j} \gamma(u) \in T_{\gamma(u)} M$ are arbitrary tangent vectors and $(\partial_{u_i u_j}^2 \gamma(u)) \perp \in (T_{\gamma(u)} M)^\perp$ is the normal component of the vector $\partial_{u_i u_j}^2 \gamma(u) \in \mathbb{R}^n$. The **mean curvature vector of** $M$ at $r(u) \in M$ is defined to be

$$ H_{r(u)} = \sum_{i,j=1}^{m} (G_{r(u)}^{-1})_{ij} b(\partial_{u_i} \gamma(u), \partial_{u_j} \gamma(u)) = \sum_{i,j=1}^{m} (G_{r(u)}^{-1})_{ij} \langle \partial_{u_i u_j}^2 \gamma(u) \rangle, $$

where $G_{r(u)}^{-1}$ denotes the inverse to the positive-definite matrix $G_{r(u)}$. The definition depends only on the embedding $M \subset V$ and not on the choice of the local parametrization. The smooth field $H_{r(u)}$, $u \in U$, of normal vectors is the **mean curvature vector field of** $M$ along the open set $r(U) \subset M$. By gluing ‘local’ definitions of the mean curvature vector field along open sets $r(U)$ from an open cover of $M$ one obtains the smooth field of normal vectors called the **mean curvature vector field of** $M$. It will be convenient to adapt the
following ‘formal’ writing of (3)

$$H_{r(u)} = \text{Tr} \left( G_{r(u)}^{-1} \right) \left( (d^2 r(u))^\perp \right),$$

(4)

where $d^2 r(u)$ is the symmetric matrix whose $(i,j)$-th entry is the vector $\frac{\partial^2}{\partial u_i \partial u_j} r(u)$ and $(\cdot)^\perp$ is applied entry-wise.

A submanifold $M \subset \mathbb{R}^n$ of the Euclidean space $(\mathbb{R}^n, \langle \cdot, \cdot \rangle)$ is called minimal if the mean curvature vector field vanishes. Thus, in order to prove minimality of $M \subset \mathbb{R}^n$ one can show that for each point $x \in M$ there is a local parametrization $r : U \to M$ with $x = r(u)$ and such that $H_{r(u)} = 0$.

Next, we state the definition of the multilinear rank of a tensor in terms of ranks of its flattening matrices.

For a tensor $T = (t_{i_1 \ldots i_d}) \in \mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_d}$ and an index $j \in [d] = \{1, \ldots, d\}$ a $j$-th flattening of $T$ is the $n_j \times \prod_{k \neq j} n_k$ matrix whose rows are indexed by $i_j \in [n_j]$, columns are indexed by $(d-1)$-tuples $(i_1, \ldots, i_{j-1}, i_{j+1}, \ldots, i_d) \in [n_1] \times \cdots \times [n_{j-1}] \times [n_{j+1}] \times \cdots \times [n_d]$ (ordered in an arbitrary but a priori fixed way) and its $(i_j, (i_1, \ldots, i_{j-1}, i_{j+1}, \ldots, i_d))$-th entry equals $t_{i_1 \ldots i_{j-1} i_j i_{j+1} \ldots i_d}$. It is well known that $T$ has multilinear rank $r = (r_1, \ldots, r_d)$ if and only if for $j \in [d]$ the $j$-th flattening matrix of $T$ has rank $r_j$, see [2, Theorem 6.13]. Finally, note that the dot products between the rows of the $j$-th flattening of $T$ with indices $\lambda, \lambda' \in [n_j]$ equals

$$\sum_{k \in [d] \setminus j} \sum_{i_k = 1}^{n_k} t_{i_1 \ldots i_{j-1} \lambda j_{j+1} \ldots i_d} t_{i_1 \ldots i_{j-1} \lambda' j_{j+1} \ldots i_d}.$$

(5)

3. Proof of main theorem

In this section we prove the main Theorem 1.1. Our goal is to show that the mean curvature vector field of $T_{d,x}$ is zero at every point $T \in T_{d,x}$. First, using the fact that the Frobenius inner product (1) is invariant under orthogonal changes of bases $\mathcal{E}^1, \ldots, \mathcal{E}^d$ of the underlying Euclidean spaces $\mathbb{R}^{n_1}, \ldots, \mathbb{R}^{n_d}$, we bring $T$ into ‘the most simple’ form (7). Second, we use a transitive action of a product of orthogonal groups to design a local parametrization of the manifold $T_{d,x}$ near a point $T$. We then use this parametrization to compute the Gram matrix, and observe its block-diagonal structure with blocks that are themselves Gram matrices of the rows of the various flattenings of the tensor. After using the block structure to show the vanishing of most terms contributing to the mean curvature, we observe that the only remaining terms are zero upon projecting to the normal space, completing the proof.

Proof of Theorem 1.1: The product of orthogonal groups $O(n) = O(n_1) \times \cdots \times O(n_d)$ acts on $\mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_d}$. More explicitly, $\mathbf{g} = (g^1, \ldots, g^d) \in O(n)$ acts on the tensor $T = (t_{i_1 \ldots i_d}) \in \mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_d}$ as

$$\mathbf{g}^* T = \left( \sum_{j=1}^d \sum_{k_j = 1}^{n_j} g^1_{i_1 k_1} \cdots g^d_{i_d k_d} t_{k_1 \ldots k_d} \right),$$

(6)
preserving the Frobenius inner product (1). This action also preserves the multilinear rank and hence restricts to an action on manifolds $T_{d,r}$. In the following, we fix an orthonormal basis $\{e^j_1, \ldots, e^j_{n_j}\}$ of $\mathbb{R}^{n_j}$. In view of the above we have $H_T = 0$ if and only if $H_{g^rT} = 0$ for any $g \in O(n)$. In particular, since the orthogonal group $O(n_j)$ acts transitively on the Grassmannian $Gr(r_j, n_j)$ of $r_j$-planes in $\mathbb{R}^{n_j}$, to prove that the mean curvature vector $H_T$ at $T \in T_{d,r}$ is zero we can first assume that $T \in W_1 \otimes \cdots \otimes W_d$ with $W_j = \text{Span}\{e^j_1, \ldots, e^j_{r_j}\}$.

$$T = \sum_{j=1}^{d} \sum_{i_1, \ldots, i_{r_j}} t_{i_1, \ldots, i_{r_j}} e^j_{i_1} \otimes \cdots \otimes e^j_{i_{r_j}}. \quad (7)$$

For $1 \leq \alpha, \beta \leq n_j$ let $E^j_{\alpha\beta}$ be the $(\alpha, \beta)$-th matrix unit of size $n_j \times n_j$,

$$\left( E^j_{\alpha\beta} \right)_{\alpha' \beta'} = \begin{cases} 1 & \text{if } \alpha = \alpha' \text{ and } \beta = \beta' \\ 0 & \text{otherwise} \end{cases},$$

and let $L^j_{\alpha\beta} = E^j_{\beta\alpha} - E^j_{\alpha\beta}$. Skew-symmetric matrices $L^j_{\alpha\beta}$, $1 \leq \alpha < \beta \leq n_j$, form a basis of the tangent space $T_{g^r}O(n_j)$ to the orthogonal group at the identity matrix $g^r \in O(n_j)$. For $L \in T_{g^r}O(n_j)$ let $u \mapsto e^{ul}$ be the one-parameter subgroup of orthogonal matrices such that $e^{0L} = g^r$ and $\frac{d}{du} e^{ul} = L e^{ul} = e^{ul} L$ and define a family of matrices in $O(n_j)$ via

$$g(u') = \prod_{\alpha=1}^{r_j} \prod_{\beta=\gamma+1}^{n_j} e^{u_{\alpha\beta} L^j_{\alpha\beta}}, \quad u' = (u^j_{\alpha\beta}), \quad (8)$$

where orthogonal matrices $e^{u_{\alpha\beta} L^j_{\alpha\beta}}$ in the product are ordered according to the order

$$(1, r_j + 1), (2, r_j + 1), \ldots, (r_j, r_j + 1), (1, r_j + 2), \ldots, (r_j, r_j + 2), \ldots, (1, n_j), \ldots, (r_j, n_j) \quad (9)$$

on the set of pairs $(\alpha, \beta)$. Note that the associated family $g(u') W_j \subset Gr(r_j, n_j)$ of $r_j$-planes contains a neighbourhood of $W_j \in Gr(r_j, n_j)$. This follows from the transitivity of the action of the orthogonal group on the Grassmannian. We will need formulas for partial derivatives of (8). By Leibniz’s rule we obtain for $1 \leq j \leq d$ and $1 \leq \lambda \leq r_j, r_j + 1 \leq \mu \leq n_j$

$$\frac{\partial g}{\partial u^j_{\lambda\mu}} (u') = \left( \prod_{\alpha=1}^{r_j} \prod_{\beta=\gamma+1}^{\omega-1} e^{u^j_{\alpha\beta} L^j_{\alpha\beta}} \right) \left( \prod_{\alpha=1}^{\lambda-1} e^{u^j_{\alpha\mu} L^j_{\alpha\mu}} \right) \times e^{u^j_{\lambda\mu} L^j_{\lambda\mu}} \left( \prod_{\alpha=\lambda+1}^{r_j} e^{u^j_{\alpha\mu} L^j_{\alpha\mu}} \right) \left( \prod_{\alpha=1}^{\mu-1} \prod_{\beta=\gamma+1}^{n_j} e^{u^j_{\alpha\beta} L^j_{\alpha\beta}} \right).$$

Evaluation of the last expression at $u' = 0$ gives

$$\frac{\partial g}{\partial u^j_{\lambda\mu}} (0) = L^j_{\lambda\mu}. \quad (10)$$
Now, the second order partial derivatives of (8) evaluated at \( u^i = 0 \) equal

\[
\frac{\partial^2 g}{\partial u^i_{\lambda\mu} \partial u^j_{\lambda'\mu'}}(0) = L^j_{\lambda,\mu} L^i_{\lambda',\mu'},
\]

where \((\lambda, \mu) \preceq (\lambda', \mu')\) with respect to the above defined order. Next we define a local parametrization of \( T_{d,r} \) around \( T = (t_{i_1...i_d}) \) by

\[
T(u^1, \ldots, u^d, S) = \sum_{j=1}^d \sum_{i_j=1}^{r_j} (t_{i_1...i_d} + s_{i_1...i_d}) g(u^1) e^{1}_{i_1} \otimes \cdots \otimes g(u^d) e^{d}_{i_d},
\]

where \( u^j \in \mathbb{R}^{r_j(n_j-r_j)}, \ j \in [d], \) and \( S = (s_{i_1...i_d}) \in \mathbb{R}^{r_1} \otimes \cdots \otimes \mathbb{R}^{r_d} \) together constitute a family of parameters of dimension \( \sum_{j=1}^d r_j(n_j - r_j) + \prod_{j=1}^d r_j = \dim(T_{d,r}). \)

Note that at \( 0 = (0, \ldots, 0, 0) = (u^1, \ldots, u^d, S) \) we have \( T(0) = T \). The first order partial derivatives of (12) evaluated at \( 0 \) equal

\[
\frac{\partial T}{\partial s_{i_1...i_d}}(0) = e^1_{i_1} \otimes \cdots \otimes e^d_{i_d}
\]

and

\[
\frac{\partial T}{\partial u^j_{\lambda,\mu}}(0) = \sum_{k=1}^d \sum_{i_k=1}^{r_k} t_{i_1...i_d} e^1_{i_1} \otimes \cdots \otimes e^{j-1}_{i_{j-1}} \otimes L^j_{\lambda,\mu} e^j_{i_j} \otimes e^{j+1}_{i_{j+1}} \otimes \cdots \otimes e^d_{i_d}
\]

\[
= \sum_{k \in [d] \setminus j} \sum_{i_k=1}^{r_k} t_{i_1...i_{j-1}i_{j+1}...i_d} e^1_{i_1} \otimes \cdots \otimes e^{j-1}_{i_{j-1}} \otimes e^j_{i_j} \otimes e^{j+1}_{i_{j+1}} \otimes \cdots \otimes e^d_{i_d},
\]

where in (14) we use (10) and the formula

\[
L^j_{\lambda,\mu} e^j_{i_j} = \begin{cases} e^j_{\mu}, & \text{if } i_j = \lambda \\ -e^j_{\lambda}, & \text{if } i_j = \mu \\ 0, & \text{otherwise} \end{cases}
\]

Note that \( \lambda \in \{1, 2, \ldots, r_j\}, \mu \in \{r_j + 1, \ldots, n_j\}, \) and the sum in (14) in the \( i_j \)-th index runs over \( i_j \in \{1, 2, \ldots, r_j\} \) so that \( i_j = \lambda \) is the only relevant case. It immediately follows from (13) and (14) that

\[
\begin{aligned}
\left\langle \frac{\partial T}{\partial s_{i_1...i_d}}(0), \frac{\partial T}{\partial s_{j_1...j_d}}(0) \right\rangle &= \delta_{i_1j_1} \cdots \delta_{i_dj_d}, \\
\left\langle \frac{\partial T}{\partial s_{i_1...i_d}}, \frac{\partial T}{\partial u^j_{\lambda,\mu}}(0) \right\rangle &= 0,
\end{aligned}
\]

\[
\begin{aligned}
\left\langle \frac{\partial T}{\partial u^j_{\lambda,\mu}}(0), \frac{\partial T}{\partial u^l_{\lambda',\mu'}}(0) \right\rangle &= \delta_{jl} \delta_{\mu\mu'} \sum_{k \in [d] \setminus j} \sum_{i_k=1}^{r_k} t_{i_1...i_{j-1}i_{j+1}...i_d} t_{i_1...i_{\lambda'-1}\lambda'_{j+1}...i_d},
\end{aligned}
\]
for every \((i_1, \ldots, i_d), (j_1, \ldots, j_d) \in [r_1] \times \cdots \times [r_d], j_i' \in [r_i], (\lambda, \mu) \in [r_j] \times \{r_j + 1, \ldots, n_j\}\) and \((\lambda', \mu') \in [r_{j'}] \times \{r_{j'} + 1, \ldots, n_{j'}\}\). With these formulas on hand we compute the Gram matrix \(G_T = T(0)\). The rows and columns of this symmetric \(\dim(T_{d,r}) \times \dim(T_{d,r})\) matrix are indexed by the local coordinates in the parametrization (12), and each entry of \(G_T\) is the inner product of the corresponding partial derivatives of (12) evaluated at \((u^1, \ldots, u^d, S) = 0\). First, formulas (15), (16) and (17) imply that \(G_T\) has a block diagonal structure, with the \(d+1\) many blocks given as

\[
G_T = \begin{pmatrix}
G_{\frac{\partial T}{\partial u^j}(0)} & 0 & \cdots & 0 & 0 \\
0 & G_{\frac{\partial T}{\partial u^j}(0)} & \ddots & \cdots & \vdots \\
\vdots & \ddots & \ddots & \cdots & \vdots \\
0 & \cdots & \cdots & G_{\frac{\partial T}{\partial u^j}(0)} & 0 \\
0 & \cdots & \cdots & 0 & G_{\frac{\partial T}{\partial S}(0)}
\end{pmatrix}.
\]

Entries of the block \(G_{\frac{\partial T}{\partial u^j}(0)}\), \(j \in [d]\), are inner products (17) with \(j = j'\), whereas the last block \(G_{\frac{\partial T}{\partial S}(0)}\) of inner products (15) is the \((\prod_{j=1}^{d} r_j \times \prod_{j=1}^{d} r_j)\) identity matrix. If we order the rows and columns of the block \(G_{\frac{\partial T}{\partial u^j}(0)}\) according to (9), then (17) with \(j = j'\) implies that \(G_{\frac{\partial T}{\partial u^j}(0)}\) has a further block structure,

\[
G_{\frac{\partial T}{\partial u^j}(0)} = \begin{pmatrix}
A_j & 0 & \cdots & 0 \\
0 & A_j & \ddots & \vdots \\
\vdots & \ddots & \ddots & 0 \\
0 & \cdots & 0 & A_j
\end{pmatrix}.
\]

More precisely, \(G_{\frac{\partial T}{\partial u^j}(0)}\) is a block diagonal \((r_j(n_j - r_j) \times r_j(n_j - r_j))\) matrix, with \(n_j - r_j\) identical blocks \(A_j\). It follows from (17) with \(j = j'\), \(\mu = \mu'\) and (5) that the \(r_j \times r_j\) matrix \(A_j\) is the Gram matrix of the rows of the \(j\)-th flattening of the tensor \(T\). Since the \(j\)-th entry \(r_j\) in the multilinear rank \(r = (r_1, \ldots, r_j, \ldots, r_d)\) of \(T \in T_{d,r}\) equals the rank of the \(j\)-th flattening (see Section 2), the matrix \(A_j\) and hence \(G_{\frac{\partial T}{\partial u^j}(0)}\) are nonsingular. Repeating the same argument for every block shows that also \(G_T\) is nonsingular and hence vectors (13) and (14) form a basis of the tangent space of \(T_{d,r}\) at \(T\).

The block diagonal structure of \(G_T\) and of \(G_{\frac{\partial T}{\partial u^j}(0)}\), \(j \in [d]\), shows that the inverse matrix \(G_T^{-1}\) is also block diagonal, with blocks of the same size. In particular we have

\[(G_T^{-1})_{\alpha \beta} = 0,
\]

in the following cases:

- \(\alpha\) corresponds to a parameter \(s_{i_1 \ldots i_d}\) and \(\beta\) corresponds to a parameter \(u^j_{\lambda, \mu'}\).
• $\alpha$ corresponds to a parameter $u^j_{\lambda, \mu}$ and $\beta$ corresponds to a parameter $u^j_{\lambda', \mu'}$, with $j \neq j'$ or with $j = j'$ but $\mu \neq \mu'$.

Next we concentrate on the computation of the mean curvature vector at the point $T$. By (4) we need to show that

$$H_T = \text{Tr} \left( G_T^{-1} \cdot (d^2 T(0))^\perp \right) = 0,$$

(18)

where $(d^2 T(0))^\perp$ is the $\dim(T_d, \mathbb{R}) \times \dim(T_d, \mathbb{R})$ matrix whose entries are normal components of second order partial derivatives of the parametrization (12) evaluated at $0 = (u^1, \ldots, u^d, S)$. Taking into account the above discovered structure of $G_T^{-1}$, we note that the only second order derivatives of (12) that are relevant for the computation of $H_T$ are

$$\frac{\partial^2 T}{\partial s_{i_1} \cdots s_{i_d} \partial s_{j_1} \cdots s_{j_d}}(0) = 0, \quad (i_1, \ldots, i_d), \ (j_1, \ldots, j_d) \in [r_1] \times \cdots \times [r_d],$$

and

$$\frac{\partial^2 T}{\partial u^j_{\lambda, \mu} \partial u^j_{\lambda', \mu}}(0) = \sum_{k=1}^{r_k} \sum_{i_k=1}^{d} \sum_{i_1=1}^{d} t_{i_1 \cdots i_d} e^{i_1}_{i_1} \otimes \cdots \otimes e^{i_{j-1}}_{i_{j-1}} \otimes e^{j_1}_{i_{j_1}} \otimes \cdots \otimes e^{j_d}_{i_d},$$

(19)

where $j \in [d]$, $\lambda, \lambda' \in [r_j]$ and $\mu \in [r_j + 1, \ldots, n_j]$. Using (11) we obtain

$$\frac{\partial g}{\partial u^j_{\lambda, \mu} \partial u^j_{\lambda', \mu}}(0) e^j_{i_j} = \left( -\delta_{\lambda, \lambda'} E^{j, \mu}_{\lambda', \mu} - E^{j, \mu}_{\lambda, \mu} \right) e^j_{i_j} = -\delta_{i_j, \lambda'} e^j_{\lambda}$$

and applying it to (19) gives

$$\frac{\partial^2 T}{\partial u^j_{\lambda, \mu} \partial u^j_{\lambda', \mu}}(0) = - \sum_{k \in [d] \setminus \{i_k\}} \sum_{i_k=1}^{d} t_{i_1 \cdots i_k \cdots i_d} e^{i_1}_{i_1} \otimes \cdots \otimes e^{i_{j-1}}_{i_{j-1}} \otimes e^{j_1}_{i_{j_1}} \otimes \cdots \otimes e^{j_d}_{i_d}.$$

Although this last computation reveals a nonzero vector we observe that all the indices $(i_1, \ldots, i_{j-1}, \lambda, i_{j+1}, \ldots, i_d)$ appearing in its expression range in $[r_1] \times \cdots \times [r_j] \times \cdots \times [r_d]$. This implies that (19) lies in the tangent space $T_T T_d, \mathbb{R}$, and hence its normal component $\left( -\frac{\partial^2 T}{\partial u^j_{\lambda, \mu} \partial u^j_{\lambda', \mu}}(0) \right)^\perp$ is zero. Thus, (18) holds and this completes the proof.

4. Linear optimization over the Segre variety

In this section we consider the Segre variety, i.e. the manifold $S_d = \mathcal{T}_{d,(1, \ldots, 1)}$ of tensors of order $d$ and multilinear rank $(1, \ldots, 1)$. Our goal is to prove Theorem 1.2. Let $T \in S_d$ and let $P \subset S_d$ be any affine hyperplane containing the tangent space to $S_d$ at $T$. We will show that in any neighbourhood of $T \in S_d$ there are rank-one tensors lying on both sides of $P$. To do so, we construct two curves $\gamma^+, \gamma^- : (-\varepsilon, \varepsilon) \to S_d$ passing through $T = \gamma^+(0) = \gamma^-(0)$ and such that $\pm\langle \gamma^+(u) - T, E \rangle > 0$ for $0 < u < \varepsilon$. The last condition simply means that the rank-one tensors $\gamma^+(u)$ and $\gamma^-(u)$ lie in the opposite open half-spaces formed by the
affine hyperplane $P_{\vec{\ell}}$. Note that the restriction of the action (6) of $O(n) = O(n_1) \times \cdots \times O(n_d)$ to $S_d$ is transitive. This together with the fact that (6) preserves the inner product (1) imply that it is enough to prove the statement when $T = e_1^1 \otimes \cdots \otimes e_1^d \in S_d$.

Before we proceed with the construction of curves $y^+$ and $y^-$, let us investigate the structure of the normal space $NTS_d$ to the Segre variety $S_d$ at $T = e_1^1 \otimes \cdots \otimes e_1^d \in S_d$. We have an orthogonal direct sum decomposition, which we denote by $\oplus^+$:

$$NTS_d = N_2 \oplus^+ N_3 \oplus^+ \cdots \oplus^+ N_d,$$

(20)

where

$$N_k = \text{Span} \{ e_1^1 \otimes \cdots \otimes w_{i_1} \otimes \cdots \otimes w_{i_k} \otimes \cdots \otimes e_1^d : 1 \leq i_1 < \cdots < i_k \leq d, w_{i_j} \in (e_1^{i_j})^\perp \}.$$ 

With these notations we may also write

$$N_0 = \text{Span} \{ e_1^1 \otimes \cdots \otimes e_1^d \} = \text{Span} \{ T \}$$

and

$$N_1 = \text{Span} \{ e_1^1 \otimes \cdots \otimes e_1^{i-1} \otimes w_i \otimes e_1^{i+1} \otimes \cdots \otimes e_1^d : 1 \leq i \leq d, w_i \in (e_1^i)^\perp \},$$

where $N_0 + N_1 = TTS_d$ is the tangent space to $S_d$ at $T$. It is straightforward to check that the subspaces $N_k$, $k = 0, \ldots, d$, are pairwise orthogonal with respect to the inner product (1).

By assumption the vector $\vec{\ell}$ belongs to the normal space $NTS_d$. Let $k \geq 2$ be the smallest natural number such that $\vec{\ell}$ has nonzero component in $N_k$. For example, if $k = 2$, the vector $\vec{\ell}$ has nonzero coefficient in front of one of the basic tensors $e_1^1 \otimes \cdots \otimes e_1^{i_1} \otimes \cdots \otimes e_1^{i_2} \otimes \cdots \otimes e_1$ for $\alpha_1, \alpha_2 > 1$ and some indices $1 \leq i_1 < i_2 \leq d$. We first prove an auxiliary lemma.

**Lemma 4.1:** Let $\vec{\ell}$ have nonzero coefficient $c(\vec{\ell}, \alpha_1, \ldots, \alpha_k) \neq 0$ in front of the basic tensor $e_1^{i_1} \otimes \cdots \otimes e_1^{i_k} \otimes e_1^{k+1} \otimes \cdots \otimes e_1^d \in N_k$, where $\alpha_1, \ldots, \alpha_k > 1$. Then the curve

$$\gamma = \{(e^u_{\alpha_1} e_1^1) \otimes \cdots \otimes (e^u_{\alpha_k} e_1^k) \otimes e_1^{k+1} \otimes \cdots \otimes e_1^d : u \in (-\varepsilon, \varepsilon) \} \subset S_d$$

(21)

satisfies $\langle \gamma^{(j)}(0), \vec{\ell} \rangle = 0$ for all $j \in \{0, 1, \ldots, k - 1\}$ and $\langle \gamma^{(k)}(0), \vec{\ell} \rangle = k! c(\vec{\ell}, \alpha_1, \ldots, \alpha_k)$.

**Proof:** This follows from Leibniz’ rule. Indeed we have that

$$\gamma(0) = e_1^1 \otimes \cdots \otimes e_1^d = T \in N_0$$

$$\gamma'(0) = \sum_{j=1}^k e_1^1 \otimes \cdots \otimes e_{\alpha_j}^j \otimes \cdots \otimes e_1^d \in N_1$$

$$\vdots$$

$$\gamma^{(k)}(0) = W + k! e_{\alpha_1}^1 \otimes \cdots \otimes e_{\alpha_k}^k \otimes e_1^{k+1} \otimes \cdots \otimes e_1^d,$$
with \( W \in N_0 \oplus N_1 \oplus \cdots \oplus N_{k-1} \). The condition \( \vec{\ell} \in N_k \), orthogonality of \( N_0, N_1, \ldots, N_{k-1}, N_k \) and orthogonality of basic tensors imply the claim.

We now prove the main result of this section.

**Proof of Theorem 1.2:** Let \( \vec{\ell} \in N_k \oplus \cdots \oplus N_d \), \( k \geq 2 \), and \( \gamma : (-\varepsilon, \varepsilon) \to \mathbb{R} \) be as in Lemma 4.1. The Taylor expansion

\[
\gamma(u) = \gamma(0) + \gamma'(0)u + \frac{\gamma''(0)}{2!}u^2 + \cdots + \frac{\gamma^{(k)}(0)}{(k-1)!}u^{k-1} + \frac{\gamma^{(k)}(0)}{k!}u^k + \cdots
\]

and Lemma 4.1 imply that the (local) position of \( \gamma(u) \) (for small \( u > 0 \)) with respect to \( P_{\vec{\ell}} = \{ T + V : \langle \vec{\ell}, V \rangle = 0 \} \) is determined by the sign of \( \langle \gamma^{(k)}(0), \vec{\ell} \rangle = k! \cdot a(\vec{\ell}, \alpha_1, \ldots, \alpha_k) \). If \( a(\vec{\ell}, \alpha_1, \ldots, \alpha_k) > 0 \) define \( \gamma^+ \) to be \( \gamma \), while if \( a(\vec{\ell}, \alpha_1, \ldots, \alpha_k) < 0 \) set \( \gamma^- = \gamma \). The curve \( \gamma^- \) (respectively \( \gamma^+ \)) is then defined to be

\[
\vec{\gamma} = \{ (e^{uL^{i_1}_{\alpha_1}} e_1^1) \otimes (e^{uL^{i_2}_{\alpha_2}} e_1^2) \otimes \cdots \otimes (e^{uL^{i_k}_{\alpha_k}} e_1^k) \otimes e_1^{k+1} \otimes \cdots \otimes e_1^d : u \in (-\varepsilon, \varepsilon) \} \subset S_d.
\]

Note that the single difference is in the sign in front of the parameter \( u \) in the first tensor factor. This change in the sign is needed to flip the sign of the inner product \( \langle \gamma^{(k)}(0), \vec{\ell} \rangle \) as \( \langle \gamma^{(j)}(0), \vec{\ell} \rangle = 0 \) for \( j = 0, 1, \ldots, k-1 \) and \( \langle \gamma^{(k)}(0), \vec{\ell} \rangle = -k! \cdot a(\vec{\ell}, \alpha_1, \ldots, \alpha_k) \). This computation follows exactly the same path as the proof of Lemma 4.1. In both cases it follows from the definition of \( \gamma^\pm \) and Taylor expansion that \( \pm \langle \gamma^\pm(u) - T, \vec{\ell} \rangle > 0 \) for small enough \( u > 0 \).

In general, if \( \vec{\ell} \in N_k \oplus \cdots \oplus N_d \) has a non-zero coefficient in front of the basic tensor \( e_1^1 \otimes \cdots \otimes e_1^{i_1} \otimes \cdots \otimes e_1^{i_k} \otimes \cdots \otimes e_1^d \in N_k \) with \( 1 \leq i_1 < \cdots < i_k \leq d \) and \( \alpha_1, \ldots, \alpha_k > 1 \), one has to consider the curve

\[
\gamma = \{ e_1^1 \otimes \cdots \otimes (e^{uL^{i_1}_{\alpha_1}} e_1^{i_1}) \otimes \cdots \otimes (e^{uL^{i_k}_{\alpha_k}} e_1^{i_k}) \otimes \cdots \otimes e_1^d : u \in (-\varepsilon, \varepsilon) \} \subset S_d
\]

instead of (21). All other steps are identical.

The above proves that there are tensors in an arbitrarily small neighborhood of \( T \in S_d \) that lie on either side of the affine hyperplane \( P_{\vec{\ell}} \). In particular, the restriction of a linear functional to \( S_d \) attains no local minima or maxima.

**Remark 4.2:** Let \( M \subseteq \mathbb{R}^n \) be a minimal submanifold such that for every point \( x \in M \) and every unit normal vector \( \vec{\ell} \in N_x M \) the second fundamental form

\[
u, \nu \in T_x M \rightarrow \langle b(u, \nu), \vec{\ell} \rangle
\]

with respect to \( \vec{\ell} \) is non-degenerate. Because of minimality of \( M \), (22) has to have a negative and a positive eigenvalue. Geometrically this means that no small neighbourhood of \( x \in M \) is completely contained in one of the half-spaces \( P_{\vec{\ell}}^\pm = \{ x + v : \pm \langle \vec{\ell}, v \rangle \geq 0 \} \). In particular, no linear functional \( L_{\vec{\ell}}(x) = \langle \vec{\ell}, x \rangle \) attains a local minimum or a local maximum on \( M \). Note however that the Segre variety \( S_d \) does not satisfy the above condition on the second
fundamental form. Indeed, one can see that all second order derivatives of the parametrization (12) evaluated at \( T = T(0) \) lie in \( N_0 \oplus N_1 \oplus N_2 \). Thus, for all \( \bar{\ell} \in N_k \) with \( k > 2 \) (see the decomposition (20)) the bilinear form (22) is trivial. That is why, in Theorem 1.2 we have to use a neater argument to prove nonexistence of local extrema of linear functionals on \( S_d \).

We now derive Corollary 1.3, which asserts that (when nonconstant) the restriction of a linear functional to the affine-linear slice \( S_d \cap A \) of the Segre variety has no local minima or maxima as well.

Figure 1. Mean curvature vector field on \( \mathcal{M} \), the independence model for two binary discrete random variables.
Proof of Corollary 1.3: Let \( A \) be the affine hyperplane \( \{ T \in \mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_d} : \langle T, \bar{a} \rangle = c \} \). Observe first that the Segre variety \( S_d \) is conical, i.e. \( \lambda T \in S_d \) for any \( T \in S_d \) and any \( \lambda \in \mathbb{R} \setminus \{0\} \). Assume that \( T \in S_d \cap A \) is a local minimum (maximum) for the restriction of a linear functional \( L_{\bar{\ell}}(V) = \langle \bar{\ell}, V \rangle \) on \( S_d \cap A \). Since the restriction of \( L_{\bar{\ell}} \) to \( A \) is assumed to be nonconstant, we have in particular that \( \bar{\ell} \) is not proportional to \( \bar{a} \). Then \( \bar{\ell}, \bar{a} \) span a two-dimensional subspace, which must intersect any hyperplane. Therefore there exists \( \mu \in \mathbb{R} \) with \( \bar{v} = \bar{\ell} + \mu \bar{a} \) orthogonal to \( T \). If \( A \) is a linear hyperplane \( (c = 0) \), then \( \langle \bar{\ell}, T \rangle = 0 \) holds automatically because in this case \( S_d \cap A \) remains conical and \( T \in T_T(S_d \cap A) \). Note that the coefficient of \( \bar{\ell} \) can be assumed equal to 1 since any rescaling of \( \bar{\ell} \) would produce another linear functional which also has \( T \) as a local minimum (maximum) on \( S_d \cap A \). The condition \( L_{\bar{v}}(T) = 0 \) implies that \( T \in S_d \cap A \) is a local minimum (maximum) of \( L_{\bar{v}} \) restricted to \( S_d \). This contradicts to Theorem 1.2. ■

Observe that while Corollary 1.3 shows that the absence of local minima and maxima is preserved under affine hyperplane sections, the same is not true for minimality. Figure 1 depicts the mean curvature field (blue vectors) for (the black pointed surface)

\[ S_2 \cap A, A = \{ S = (s_{ij})_{i,j=1,2} \in \mathbb{R}^2 \otimes \mathbb{R}^2 : s_{11} + s_{12} + s_{21} + s_{22} = 1 \} \]

Since this vector field is evidently nonzero, this surface \( S_2 \cap A \) is not minimal. If we restrict ourselves to tensors with nonnegative entries we obtain a statistical independence model, as explained in the introduction. This is precisely the portion of \( S_2 \cap A \) captured in Figure 1. We make the code to compute the mean curvature for this example available at https://mathrepo.mis.mpg.de/tensorminimality.

Disclosure statement
No potential conflict of interest was reported by the author(s).

ORCID
Alexander Heaton http://orcid.org/0000-0003-0422-3474

References
[1] Lim L-H. Tensors in computations. Acta Numer. 2021;30:555–764.
[2] Hackbusch W. Tensor spaces and numerical tensor calculus. Vol. 42, Springer, Berlin, Heidelberg; 2012 Jan.
[3] Hitchcock FL. The expression of a tensor or a polyadic as a sum of products. J Math Phys. 1927;6(1–4):164–189.
[4] Uschmajew A, Vandereycken B. The geometry of algorithms using hierarchical tensors. Linear Algebra Appl. 2013;439(1):133–166.
[5] Comon P, Lim L-H, Qi Y, et al. Topology of tensor ranks. Adv Math. 2020;367:107–128.
[6] De Lathauwer L, De Moor B, Vandewalle J. A multilinear singular value decomposition. SIAM J Matrix Anal Appl. 2000;21(4):1253–1278.
[7] Tucker L.R. Some mathematical notes on three-mode factor analysis. Psychometrika. 1966;31:279–311.
[8] Vasilescu M, Alex O, Terzopoulos D. Multilinear analysis of image ensembles: tensorfaces. Proceedings of the 7th European Conference on Computer Vision-Part I, ECCV ’02; Berlin: Springer-Verlag; 2002. p. 447–460.
[9] Landsberg JM. Tensors: geometry and applications. Providence (RI): American Mathematical Society; 2012. (Graduate Studies in Mathematics; 128).
[10] De Lathauwer L, De Moor B, Vandewalle J. On the best rank-1 and rank-$(R_1, R_2, \ldots, R_N)$ approximation of higher-order tensors. SIAM J Matrix Anal Appl. 2000;21(4):1324–1342.
[11] Eckart C, Young GM. The approximation of one matrix by another of lower rank. Psychometrika. 1936;1:211–218.
[12] de Silva V, Lim L-H. Tensor rank and the ill-posedness of the best low-rank approximation problem. SIAM J Matrix Anal Appl. 2008;30(3):1084–1127.
[13] Grasedyck L, Kressner D, Tobler C. A literature survey of low-rank tensor approximation techniques. GAMM Mitteilungen. 2013;36(1):53–78.
[14] Absil P-A, Mahony R, Sepulchre R. Optimization algorithms on matrix manifolds. Princeton (NJ): Princeton Univ. Press; 2008.
[15] Edelman A, Arias TA, Smith ST. The geometry of algorithms with orthogonality constraints. SIAM J Matrix Anal Appl. 1999;20(2):303–353.
[16] Kressner D, Steinlechner M, Vandereycken B. Low-rank tensor completion by riemannian optimization. BIT Numer Math. 2014;54(2):447–468.
[17] Da Silva C, Herrmann FJ. Optimization on the hierarchical Tucker manifold – applications to tensor completion. Linear Algebra Appl. 2015;481:131–173.
[18] Eldén L, Savas B. A Newton-Grassmann method for computing the best multilinear rank-(r1; r2; r3) approximation of a tensor. SIAM J Matrix Anal Appl. 2009;31(2):248–271.
[19] Savas B, Lim L-H. Quasi-Newton methods on Grassmannians and multilinear approximations of tensors. SIAM J Sci Comput. 2010;32(6):3352–3393.
[20] Koch O, Lubich C. Dynamical tensor approximation. SIAM J Matrix Anal Appl. 2010;31(5):2360–2375.
[21] Bordemann M, Choe J, Hoppe J. The minimality of determinantal varieties. J für die Reine und Angew Math. 2021;2021(773):153–164.
[22] Kozhasov Kh. On minimality of determinantal varieties. Linear Algebra Appl. 2021;626:56–78.
[23] Tkachev V. Minimal cubic cones via clifford algebras. Complex Anal Oper Theory. 2010;4:685–700.
[24] Ö. Çelik T, Jamneshan A, Montúfar G, et al. Wasserstein distance to independence models. J Symb Comput. 2021;104:855–873.
[25] Chavel I. Riemannian geometry: a modern introduction. 2nd ed. Cambridge: Cambridge University Press; 2006. (Cambridge studies in advanced mathematics; 98).
[26] Kobayashi Sh, Nomizu K. Foundations of Differential Geometry. Vol. 2, New York, Interscience publishers; 1969.