CHEBYSHEV POLYNOMIALS AND BEST RANK-ONE APPROXIMATION RATIO[*]

ANDREI AGRACHEV[†], KHAZHGALI KOZHASOV[‡], AND ANDRÉ USCHMAJEW[§]

Abstract. We establish a new extremal property of the classical Chebyshev polynomials in the context of best rank-one approximation of tensors. We also give some necessary conditions for a tensor to be a minimizer of the ratio of spectral and Frobenius norms.

Introduction and outline. The classical Chebyshev polynomials are known to have many extremal properties. The first result was established by Chebyshev himself: he proved [3] that a univariate monic polynomial with real coefficients that least deviates from zero on the interval $[-1, 1]$ must be proportional to a Chebyshev polynomial of the first kind. Later there were further developments highlighting extremal properties of this class of univariate polynomials and its relevance for approximation theory; see [11, 17] and references therein. In this article we discover a new extremal property of Chebyshev polynomials of the first kind in the context of the theory of rank-one approximations of real tensors.

Let us define *the binary Chebyshev form of degree d* as

$$(0.1) \qquad Ч_{d,2}(x_1, x_2) = \frac{(x_1 + ix_2)^d + (x_1 - ix_2)^d}{2} = \sum_{k=0}^{[d/2]} \binom{d}{2k}(-1)^k x_1^{d-2k} x_2^{2k}.$$

Note that its restriction to the unit circle $x_1^2 + x_2^2 = 1$ can be identified with the univariate Chebyshev polynomial of the first kind $x \mapsto Ч_{d,2}(x, \sqrt{1-x^2}) = \cos(d \arccos x)$, $x \in [-1, 1]$. In Theorem 1.1 we prove that the binary form (0.1) minimizes the ratio of the uniform norm on the unit circle and the Bombieri norm among all nonzero binary forms of the given degree d.

In [18], the more general problem of minimizing the ratio of the uniform norm on the unit sphere and the Bombieri norm among all nonzero forms of a given degree d and number of variables n was considered. Equivalently, identifying a homogeneous polynomial with the symmetric tensor of its coefficients, one can formulate this problem as follows: minimize the ratio of the spectral norm and the Frobenius norm among all nonzero real symmetric n^d-tensors. In an attempt to attack this problem we define the family of homogeneous n-ary forms (1.2) that we call *Chebyshev forms* $Ч_{d,n}$.

Besides solving the above problem for the case of binary forms in Theorem 1.1, we solve it in the case of cubic ternary forms ($d = 3$, $n = 3$) in Theorem 1.2. This latter result in fact follows from a more general result that we obtain in Theorem 1.5: the maximal orthogonal rank of a real $(3, 3, 3)$-tensor is 7. This in particular implies that the minimum value of the ratio of the spectral norm and the Frobenius norm of a nonzero $(3, 3, 3)$-tensor is $1/\sqrt{7}$ and hence gives an affirmative answer to a conjecture in [14]. Since the spectral norm of a tensor measures its relative distance to the set of rank-one tensors (see section 2.2), yet another way to interpret this result is that the symmetric tensor associated to the Chebyshev form Ч$_{3,3}$ achieves the maximum possible relative distance to the set of all rank-one $(3, 3, 3)$-tensors.

In Theorem 1.10 we show that if a tensor minimizes the ratio of the spectral and the Frobenius norms, then it lies in the space spanned by its best rank-one approximations. In Theorem 1.11 we prove an analogous result for symmetric tensors or, equivalently, homogeneous forms: if a form minimizes the ratio of the uniform norm on the unit sphere and the Bombieri norm, then it lies in the space spanned by rank-one forms defined by global extrema of the restriction of the form to the unit sphere. These two results imply lower bounds on the number of best rank-one approximations for those tensors (respectively, on the number of global extrema of homogeneous forms) that minimize the ratio of norms; see Corollary 1.12.

In the next section we state all our results in detail. The results are proved in section 3. Section 2 contains some necessary preliminaries and auxiliary results.

1. Main results. In this section we state our main results. They are all closely related but can be grouped into somewhat different directions.

1.1. Chebyshev forms and their extremal property. In the following $P_{d,n}$ denotes the space of real n-ary forms of degree d (real homogeneous polynomials of degree d in n variables), and $\|x\| = \sqrt{x_1^2 + \cdots + x_n^2}$ is the Euclidean norm on $\mathbb{R}^n$. For a form p we denote by

$$\|p\|_\infty = \max_{\|x\|=1} |p(x)|$$

the *uniform norm* of its restriction to the unit sphere.

Every form $p \in P_{d,n}$ has a standard representation in the basis of monomials: $p(x) = \sum_{|\alpha|=d} c_\alpha x^\alpha \in P_{d,n}$, where $\alpha = (\alpha_1, \ldots, \alpha_n) \in \{0, 1, \ldots, d\}^n$ is a multi-index of length $|\alpha| = \alpha_1 + \cdots + \alpha_n = d$ and $x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$. The *Bombieri norm* [2] of p is defined as

$$\|p\|_B = \sqrt{\sum_{|\alpha|=d} \binom{d}{\alpha}^{-1} |c_\alpha|^2},$$

where $\binom{d}{\alpha} = \frac{d!}{\alpha_1! \ldots \alpha_n!}$ is the multinomial coefficient.

The *conformal orthogonal group* $CO(n) = \mathbb{R}_+ \times O(n)$ acts on the space $P_{d,n}$ of real forms as follows:

$$g = (s, \rho) \in CO(n), \ p \in P_{d,n} \ \mapsto \ g^*p \in P_{d,n}, \quad (g^*p)(x) = sp(\rho^{-1}x).$$

Note that both the uniform norm and the Bombieri norm are invariant under the subgroup $O(n)$ of orthogonal transformations and their ratio is invariant under the full group $CO(n)$; see section 2.2.

In [18], Qi asked about the smallest possible ratio $\|p\|_\infty / \|p\|_B$ that the two norms can attain in a space $P_{d,n}$. In our first result we solve this problem for binary forms of any given degree and we also characterize minimizers in this case.

THEOREM 1.1. *For any nonzero $p \in P_{d,2}$ it holds that*

$$
(1.1) \qquad \frac{\|p\|_\infty}{\|p\|_B} \geq \frac{\|\text{Ч}_{d,2}\|_\infty}{\|\text{Ч}_{d,2}\|_B} = \frac{1}{\sqrt{2^{d-1}}}.
$$

When $d = 0, 1$ one has equality in (1.1) for any $p \in P_{d,2}$, when $d = 2$ equality holds if and only if $p = \pm g^*(x_1^2 + x_2^2)$ or $p = g^* \text{Ч}_2 = g^*(x_1^2 - x_2^2)$, where $g \in CO(2)$. When $d \geq 3$ equality holds if and only if $p = g^* \text{Ч}_{d,2}$, $g \in CO(2)$.

For any $d \geq 0$ and $n \geq 2$ we define the *n-ary Chebyshev form of degree d* as

$$
(1.2) \qquad \text{Ч}_{d,n}(x_1, \ldots, x_n) = \sum_{k=0}^{[d/2]} \binom{d}{2k} (-1)^k x_1^{d-2k} (x_2^2 + \cdots + x_n^2)^k.
$$

Note that the forms $\text{Ч}_{d,n}$ are invariant under orthogonal transformations of $\mathbb{R}^n$ that preserve the point $(1, 0, \ldots, 0)$ and for any vector $v = (v_2, \ldots, v_n) \in \mathbb{R}^{n-1}$ of unit length one has that $\text{Ч}_{d,n}(x_1, v_2 x_2, \ldots, v_n x_2) = \text{Ч}_{d,2}(x_1, x_2)$ is the binary Chebyshev form (0.1). In this work we are particularly concerned with cubic Chebyshev forms

$$
(1.3) \qquad \text{Ч}_{3,n}(x_1, \ldots, x_n) = x_1^3 - 3x_1(x_2^2 + \cdots + x_n^2).
$$

It is an easy calculation that

$$
(1.4) \qquad \|\text{Ч}_{d,n}\|_\infty = 1, \quad \|\text{Ч}_{d,n}\|_B^2 = \sum_{k=0}^{[d/2]} \binom{d}{2k} \sum_{\substack{\beta = (\beta_1, \ldots, \beta_{n-1}), \\ |\beta| = k}} \binom{k}{\beta}^2 \binom{2k}{2\beta}^{-1},
$$

where $2\beta = (2\beta_1, \ldots, 2\beta_{n-1})$, and, in particular,

$$
(1.5) \qquad \|\text{Ч}_{3,n}\|_B^2 = 3n - 2.
$$

In the case $d = 2$ of quadratic forms one can easily determine the minimal ratio $\|p\|_\infty / \|p\|_B$ by passing to the ratio of spectral and Frobenius norms of symmetric matrices. Specifically,

$$
\frac{\|p\|_\infty}{\|p\|_B} \geq \frac{1}{\sqrt{n}}, \quad p \in P_{2,n},
$$

with equality only for quadratic forms $p = g^*(\pm x_1^2 \pm \cdots \pm x_n^2)$, where $g \in CO(n)$. These forms correspond to multiples of symmetric orthogonal matrices. Note that among these extremal quadratic forms there is the Chebyshev quadric $\text{Ч}_{2,n}(x) = x_1^2 - x_2^2 - \cdots - x_n^2$ which is classically known as *the Lorentz quadric*. This fact for $d = 2$ together with Theorem 1.1 might suggest thinking that Chebyshev forms $\text{Ч}_{d,n}$ also minimize the ratio of uniform and Bombieri norm in $P_{d,n}$ for $d \geq 3$ and $n \geq 3$. We show that it is indeed the case for the first "nontrivial" situation $d = 3$, $n = 3$ of ternary cubics.

THEOREM 1.2. *Let $p \in P_{3,3}$ be a nonzero ternary cubic form. Then*

$$
\frac{\|p\|_\infty}{\|p\|_B} \geq \frac{1}{\sqrt{7}}
$$

and equality holds if $p = g^* \text{Ч}_{3,3}$, where $g \in CO(3)$.

Theorem 1.2 is part of Corollary 1.6 further below.

However, at least for all sufficiently large n, the Chebyshev form Ч$_{3,n}$ is not a global minimizer for the norm ratio. Indeed, [16, Theorem 5.3] provides examples of symmetric $n \times n \times n$ tensors with $n = 2^m$ that yield forms $p \in P_{3,2^m}$ satisfying

$$(1.6) \qquad \frac{\|p\|_\infty}{\|p\|_B} = \left(\frac{2}{3}\right)^m = n^{\frac{\ln(2/3)}{\ln 2}} \leq n^{-0.584},$$

whereas, by (1.4) and (1.5),

$$\frac{\|Ч_{3,n}\|_\infty}{\|Ч_{3,n}\|_B} = \frac{1}{\sqrt{3n-2}}.$$

For instance, for $n = 2^{10} = 1024$, it holds that $\|Ч_{3,n}\|_\infty / \|Ч_{3,n}\|_B \geq 0.0187 > (2/3)^{10} \approx 0.0173$.

Interestingly, while not being a global minimum, one can show that Ч$_{3,n}$ is a local minimum of the ratio of the two norms on the set of nonzero cubic n-ary forms.

THEOREM 1.3. *Let $n \geq 2$. For all $p \in P_{3,n}$ in a small neighborhood of Ч$_{3,n}$ we have*

$$\frac{\|p\|_\infty}{\|p\|_B} \geq \frac{\|Ч_{3,n}\|_\infty}{\|Ч_{3,n}\|_B}.$$

1.2. Best rank-one approximation ratio, orthogonal rank, and orthogonal tensors. Let $\otimes_{j=1}^d \mathbb{R}^{n_j}$ denote the space of real $(n_1, \ldots, n_d)$-tensors, considered as $n_1 \times \cdots \times n_d$ tables $A = (a_{i_1 \ldots i_d})$ of real numbers. For two $(n_1, \ldots, n_d)$-tensors their *Frobenius inner product* is given by

$$\langle A, A' \rangle_F = \sum_{i_1, \ldots, i_d = 1}^{n_1, \ldots, n_d} a_{i_1 \ldots i_d} a'_{i_1 \ldots i_d}$$

and $\|A\|_F = \sqrt{\langle A, A \rangle_F}$ denotes the induced Frobenius norm.

The outer product $x^{(1)} \otimes \cdots \otimes x^{(d)}$ of vectors $x^{(j)} \in \mathbb{R}^{n_j}$ is an $(n_1, \ldots, n_d)$-tensor X with entries $(x_{i_1}^{(1)} \cdots x_{i_d}^{(d)})$. Nonzero tensors of this form are said to be of *rank one*, denoted $\text{rank}(X) = 1$. The *spectral norm* on $\otimes_{j=1}^d \mathbb{R}^{n_j}$ is defined as

$$(1.7) \qquad \|A\|_2 = \max_{\|x^{(1)}\| = \cdots = \|x^{(d)}\| = 1} \langle A, x^{(1)} \otimes \cdots \otimes x^{(d)} \rangle_F = \max_{\substack{\|X\|_F = 1, \\ \text{rank}(X) = 1}} \langle A, X \rangle_F,$$

where $\| \cdot \|$ denotes the standard Euclidean norm.

Given a $(n_1, \ldots, n_d)$-tensor A, a rank-one tensor $Y \in \otimes_{j=1}^d \mathbb{R}^{n_j}$ is called *a best rank-one approximation to A* if it minimizes the Frobenius distance to A from the set of rank-one tensors, that is,

$$(1.8) \qquad \|A - Y\|_F = \min_{X \in \otimes_{j=1}^d \mathbb{R}^{n_j}, \, \text{rank}(X) = 1} \|A - X\|_F.$$

The notion of *best rank-one approximation ratio* of a tensor space was introduced by Qi in [18]. For the space of $(n_1, \ldots, n_d)$-tensors it is defined as

$$(1.9) \qquad \mathcal{A}\left(\otimes_{j=1}^d \mathbb{R}^{n_j}\right) = \min_{0 \neq A \in \otimes_{j=1}^d \mathbb{R}^{n_j}} \frac{\|A\|_2}{\|A\|_F}.$$

It is the largest constant c satisfying $\|A\|_2 \geq c\|A\|_F$ for all $A \in \otimes_{j=1}^d \mathbb{R}^{n_j}$. Another interpretation is that $\mathcal{A}(\otimes_{j=1}^d \mathbb{R}^{n_j})$ is the inverse of the operator norm of the identity map from $(\otimes_{j=1}^d \mathbb{R}^{n_j}, \|\cdot\|_2)$ to $(\otimes_{j=1}^d \mathbb{R}^{n_j}, \|\cdot\|_F)$.

DEFINITION 1.4. *A nonzero tensor* $A \in \otimes_{j=1}^d \mathbb{R}^{n_j}$ *is called* extremal *if it is a minimizer in* (1.9), *that is, if it satisfies*

$$\frac{\|A\|_2}{\|A\|_F} = \mathcal{A}\left(\otimes_{j=1}^d \mathbb{R}^{n_j}\right).$$

Seen as a function of a tensor $A \in \otimes_{j=1}^d \mathbb{R}^{n_j}$, $\|A\|_F = 1$, of unit Frobenius norm, the rank-one approximation error (1.8) attains its maximum exactly at extremal tensors of unit Frobenius norm. The precise relation between (1.8) and (1.9) together with a possible application is given in (2.13) in subsection 2.2.

The space $\mathrm{Sym}^d(\mathbb{R}^n)$ of symmetric n^d-tensors consists of tensors $A = (a_{i_1 \ldots i_d})$ in $\otimes_{j=1}^d \mathbb{R}^n$ that satisfy $a_{i_{\sigma_1} \ldots i_{\sigma_d}} = a_{i_1 \ldots i_d}$ for any permutation σ on d elements. This space is isomorphic to the space $P_{d,n}$ of homogeneous forms as explained in subsection 2.1. Under this isomorphism Frobenius and spectral norms of a symmetric tensor correspond to Bombieri norm and uniform norm, respectively. *The best rank-one approximation ratio* $\mathcal{A}(\mathrm{Sym}^d(\mathbb{R}^n))$ *of the space of symmetric tensors is defined by* replacing $\otimes_{j=1}^d \mathbb{R}^{n_j}$ with $\mathrm{Sym}^d(\mathbb{R}^n)$ in (1.9) and is equal to the minimum ratio between the uniform and the Bombieri norms of a nonzero form in $P_{d,n}$. In this context it is important to note that the definition of the spectral norm of a symmetric tensor does not change if the maximum in (1.7) is taken over symmetric rank-one tensors only; see subsection 2.1.

A general formula for $\mathcal{A}(\otimes_{j=1}^d \mathbb{R}^{n_j})$ or $\mathcal{A}(\mathrm{Sym}^d(\mathbb{R}^n))$ is not known except for special cases; see [15]. Determining or estimating these constants is an interesting problem on its own and may have some useful applications for rank-truncated tensor optimization methods (see section 2.2). The present work contains some new contributions with the main focus on symmetric tensors.

One always has

$$0 < \mathcal{A}\left(\otimes_{j=1}^d \mathbb{R}^{n_j}\right) \leq 1 \quad \text{and} \quad 0 < \mathcal{A}(\otimes_{j=1}^d \mathbb{R}^n) \leq \mathcal{A}(\mathrm{Sym}^d(\mathbb{R}^n)) \leq 1.$$

The asymptotic behavior of $\mathcal{A}(\otimes_{j=1}^d \mathbb{R}^n)$ is $O(1/\sqrt{n^{d-1}})$; see [7]. For $d = 3$ the currently best known upper bound valid for all n seems to be $1.5n^{\ln(2/3)/\ln 2} \leq 1.5n^{-0.584}$ and follows directly from (1.6); see [16].

Lower bounds on the best rank-one approximation ratio can be obtained from decomposition of tensors into pairwise orthogonal rank-one tensors. For $A \in \otimes_{j=1}^d \mathbb{R}^{n_j}$ let

$$(1.10) \qquad\qquad A = Y_1 + \cdots + Y_r,$$

where $Y_1, \ldots, Y_r$ are rank-one $(n_1, \ldots, n_d)$-tensors such that $\langle Y_\ell, Y_{\ell'} \rangle_F = 0$ for $\ell \neq \ell'$. The smallest possible number r that allows such a decomposition (1.10) is called the *orthogonal rank* of the tensor A [9] and will be denoted by $\mathrm{rank}_\perp(A)$. Since at least one of the terms in (1.10) has to satisfy $\langle A, Y_i \rangle_F \geq \|A\|_F^2/r$, it follows that

$$\frac{\|A\|_2}{\|A\|_F} \geq \frac{1}{\sqrt{\mathrm{rank}_\perp(A)}}$$

for all $A \in \otimes_{j=1}^{d} \mathbb{R}^{n_j}$. Thus an upper bound on the maximal orthogonal rank in a given tensor space leads to a lower bound on the best rank-one approximation ratio of that tensor space:

$$(1.11) \qquad \mathcal{A}\left(\otimes_{j=1}^{d} \mathbb{R}^{n_j}\right) \geq \frac{1}{\sqrt{\max_{\otimes_{j=1}^{d} \mathbb{R}^{n_j}} \operatorname{rank}_\perp(A)}}.$$

It appears that for all known values of $\mathcal{A}(\otimes_{j=1}^{d} \mathbb{R}^{n_j})$ this is actually an equality [13, 15].

The values for $\mathcal{A}(\mathbb{R}^{n_1} \otimes \mathbb{R}^{n_2} \otimes \mathbb{R}^{n_3})$ have been determined in [14] for all combinations $n_1, n_2, n_3 \leq 4$, except for $(3, 3, 3)$-tensors. In the present work we are able to settle this remaining case, by determining the maximum possible orthogonal rank of a $(3, 3, 3)$-tensor.

THEOREM 1.5. *The maximal orthogonal rank of a $(3, 3, 3)$-tensor is seven.*

In [14] it has been shown that $1/\sqrt{7}$ is an upper bound for $\mathcal{A}(\mathbb{R}^3 \otimes \mathbb{R}^3 \otimes \mathbb{R}^3)$ and conjectured that it is actually the exact value. Due to (1.11), Theorem 1.5 shows that $1/\sqrt{7}$ is also a lower bound and hence proves this conjecture. On the other hand, we see from (1.4) and (1.5) that the minimal ratio $1/\sqrt{7}$ can be achieved by symmetric $(3, 3, 3)$-tensors, in particular by the ones associated with the Chebyshev form $Ч_{3,3}$. Since the spaces $\operatorname{Sym}^3(\mathbb{R}^3)$ and $P_{3,3}$ are isometric (with respect to the both norms), Theorem 1.2 is therefore part of the following corollary of Theorem 1.5.

COROLLARY 1.6. *We have*

$$\mathcal{A}(\mathbb{R}^3 \otimes \mathbb{R}^3 \otimes \mathbb{R}^3) = \mathcal{A}(\operatorname{Sym}^3(\mathbb{R}^3)) = \frac{1}{\sqrt{\max_{A \in \mathbb{R}^3 \otimes \mathbb{R}^3 \otimes \mathbb{R}^3} \operatorname{rank}_\perp(A)}} = \frac{1}{\sqrt{7}}$$

and the symmetric tensor corresponding to the Chebyshev cubic $Ч_{3,3}$ is extremal.

Assume now that $n_1 \leq \cdots \leq n_d$. Then it is not difficult to show that the orthogonal rank of an $(n_1, \ldots, n_d)$-tensor is not larger than $n_1 \cdots n_{d-1}$. It follows from (1.11) that

$$(1.12) \qquad \mathcal{A}\left(\otimes_{j=1}^{d} \mathbb{R}^{n_j}\right) \geq \frac{1}{\sqrt{n_1 \cdots n_{d-1}}}, \quad n_1 \leq \cdots \leq n_d.$$

In [15] the concept of an *orthogonal tensor* is defined by the property that its contraction along the first $d-1$ modes (assuming n_d is the largest dimension) with any $d-1$ vectors of unit length results in a vector of unit length. It is then shown that equality in (1.12) is attained if and only if the space contains orthogonal tensors and only those are then the extremal ones. Moreover, for n^d-tensors this is the case if and only if $n = 1, 2, 4, 8$. Therefore, Theorem 1.1 in particular shows that

$$\mathcal{A}(\operatorname{Sym}^d(\mathbb{R}^2)) = \frac{1}{\sqrt{2^{d-1}}} = \mathcal{A}(\otimes_{j=1}^{d} \mathbb{R}^2),$$

and since the symmetric tensors associated to Chebyshev forms attain these constants, they are orthogonal in the sense of [15]. In light of Corollary 1.6 one hence may wonder whether $\mathcal{A}(\operatorname{Sym}^d(\mathbb{R}^n))$ equals $\mathcal{A}(\otimes_{j=1}^{d} \mathbb{R}^n)$ in general, or at least in the case $d = 3$. Note that this is true for matrices. In general, the answer to this question is, however, negative. In the cases $n = 4$ and $n = 8$ it would imply the existence of symmetric orthogonal tensors, which we show is not possible.

PROPOSITION 1.7. *If $A \in \mathrm{Sym}^d(\mathbb{R}^n)$ is an orthogonal symmetric tensor of order $d \geq 3$, then $n = 1$ or $n = 2$. For $n = 2$ the only such tensors are the ones associated to rotated Chebyshev forms $p = \rho^* Ч_{d,2}$, $\rho \in O(2)$, that is, are of the form $(\rho, \ldots, \rho) \cdot A$ (see (2.4)) with A given by (2.7).*

COROLLARY 1.8. *For $d \geq 3$ we have*

$$\mathcal{A}\left(\otimes_{j=1}^d \mathbb{R}^4\right) = \frac{1}{\sqrt{4^{d-1}}} < \mathcal{A}(\mathrm{Sym}^d(\mathbb{R}^4)) \ \text{ and } \ \mathcal{A}\left(\otimes_{j=1}^d \mathbb{R}^8\right) = \frac{1}{\sqrt{8^{d-1}}} < \mathcal{A}(\mathrm{Sym}^d(\mathbb{R}^8)).$$

The cases of 2^d- and $(3, 3, 3)$-tensors are therefore exceptional in the sense that the "nonsymmetric" best rank-one approximation ratio can be achieved by symmetric tensors.

1.3. Variational characterization and critical tensors. The problem of determining the best rank-one approximation ratio of a tensor space and finding associated extremal tensors can be seen as a constrained optimization problem for a Lipschitz function. The spectral norm $A \mapsto \|A\|_2$ is a Lipschitz function on the normed space $(\otimes_{j=1}^d \mathbb{R}^{n_j}, \| \cdot \|_F)$ (with Lipschitz constant one). The best rank-one approximation ratio $\mathcal{A}(\otimes_{j=1}^d \mathbb{R}^{n_j})$ equals the minimal value of this function on the unit sphere $\{A \in \otimes_{j=1}^d \mathbb{R}^{n_j} : \|A\|_F = 1\}$ defined by the Frobenius norm, and extremal tensors (of unit Frobenius norm) are its global minima. Global as well as local minima of a Lipschitz function are among its critical points. The notion of a critical point of a Lipschitz function constrained to a submanifold is explained in section 2.3. It motivates the following terminology.

DEFINITION 1.9. *A nonzero tensor $A \in \otimes_{j=1}^d \mathbb{R}^{n_j}$ is* critical *if $A/\|A\|_F$ is a critical point of the restriction of the spectral norm to the Frobenius sphere, meaning that λA belongs to the generalized gradient of the spectral norm at $A/\|A\|_F$ for some $\lambda \in \mathbb{R}$.*

We can then give a characterization of critical $(n_1, \ldots, n_d)$-tensors in terms of decompositions of them into their best rank-one approximations.

THEOREM 1.10. *A nonzero tensor $A \in \otimes_{j=1}^d \mathbb{R}^{n_j}$ is critical if and only if the rescaled tensor $\|A\|_2^2/\|A\|_F^2 \, A$ can be written as a convex linear combination of some best rank-one approximations of A.*

Specifically, the theorem states that a tensor A is critical if and only if there exists a decomposition

$$(1.13) \qquad \left(\frac{\|A\|_2}{\|A\|_F}\right)^2 A = \sum_{\ell=1}^r \alpha_\ell Y_\ell, \quad \sum_{\ell=1}^r \alpha_\ell = 1, \ \alpha_1, \ldots, \alpha_r > 0,$$

where $Y_1, \ldots, Y_r$ are best rank-one approximations of A. In particular, if $A \in \otimes_{j=1}^d \mathbb{R}^{n_j}$ is an extremal tensor, then

$$(1.14) \qquad \mathcal{A}\left(\otimes_{j=1}^d \mathbb{R}^{n_j}\right)^2 A = \sum_{\ell=1}^r \alpha_\ell Y_\ell, \quad \sum_{\ell=1}^r \alpha_\ell = 1, \ \alpha_1, \ldots, \alpha_r > 0$$

for some best rank-one approximations $Y_1, \ldots, Y_r$ of A.

An analogue of Theorem 1.10 holds for symmetric tensors or, equivalently, homogeneous forms. Considering the spectral norm as a function on the space $\mathrm{Sym}^d(\mathbb{R}^n)$ only, it is again a Lipschitz function, and the best rank-one approximation ratio of $\mathrm{Sym}^d(\mathbb{R}^n)$ equals its minimum value on the Frobenius unit sphere in the space

$\mathrm{Sym}^d(\mathbb{R}^n)$ of symmetric tensors. A nonzero symmetric tensor $A \in \mathrm{Sym}^d(\mathbb{R}^n)$ is called *critical in* $\mathrm{Sym}^d(\mathbb{R}^n)$ if the normalized symmetric tensor $A/\|A\|_F$ is a critical point (see section 2.3) of the restriction of the spectral norm to the Frobenius sphere in the space $\mathrm{Sym}^d(\mathbb{R}^n)$. We also say that a form $p \in P_{d,n}$ is critical if the associated symmetric tensor is critical in $\mathrm{Sym}^d(\mathbb{R}^n)$.

THEOREM 1.11. *A nonzero tensor* $A \in \mathrm{Sym}^d(\mathbb{R}^n)$ *is critical in* $\mathrm{Sym}^d(\mathbb{R}^n)$ *if and only if the rescaled tensor* $\|A\|_2^2/\|A\|_F^2 \, A$ *can be written as a convex linear combination of some symmetric best rank-one approximations of* A. *In this case* A *is also critical in the space* $\otimes_{j=1}^d \mathbb{R}^n$.

Here the second statement follows immediately from Theorem 1.10 and the fact that a best rank-one approximation of a symmetric tensor can always be chosen to be symmetric due to Banach's result [1]; see section 2.1. However, if $A \in \mathrm{Sym}^d(\mathbb{R}^n)$ is an extremal symmetric tensor, then, by Theorem 1.11,

$$(1.15) \qquad \mathcal{A}(\mathrm{Sym}^d(\mathbb{R}^n))^2 A = \sum_{\ell=1}^r \alpha_\ell Y_\ell, \quad \sum_{\ell=1}^r \alpha_\ell = 1, \ \alpha_1, \ldots, \alpha_r > 0$$

for some symmetric best rank-one approximations $Y_1, \ldots, Y_r$ of A, and A is critical in $\otimes_{j=1}^d \mathbb{R}^{n_j}$. But in general A is not extremal in $\otimes_{j=1}^d \mathbb{R}^{n_j}$ as discussed at the end of the previous subsection.

Theorems 1.10 and 1.11 combined with Proposition 2.2 from section 2.2 imply that extremal tensors must have several best rank-one approximations.

COROLLARY 1.12. *Let* $d \geq 2$. *Then any extremal tensor in* $\otimes_{j=1}^d \mathbb{R}^n$ *has at least* n *distinct best rank-one approximations. Similarly, any extremal symmetric tensor in* $\mathrm{Sym}^d(\mathbb{R}^n)$ *has at least* n *distinct symmetric best rank-one approximations.*

Below we give an alternative characterization of critical tensors in terms of their nuclear norm. The nuclear norm of a $(n_1, \ldots, n_d)$-tensor $A \in \otimes_{j=1}^d \mathbb{R}^{n_j}$ is defined by

$$(1.16) \quad \|A\|_* = \inf \left\{ \sum_{\ell=1}^r \|Y_\ell\|_F : \ A = \sum_{\ell=1}^r Y_\ell, \ r \in \mathbb{N}, \ \mathrm{rank}(Y_\ell) = 1, \ell = 1, \ldots, r \right\}.$$

It is a result of Friedland and Lim [10] that for a symmetric tensor $A \in \mathrm{Sym}^d(\mathbb{R}^n)$ it is enough to take the infimum in (1.16) over symmetric rank-one tensors only. Hence the nuclear norm of a symmetric tensor can be defined intrinsically in the space $\mathrm{Sym}^d(\mathbb{R}^n)$. In either case, the infimum in (1.16) is attained.

Nuclear and spectral norms are dual to each other (see subsection 2.1) and for any tensor $A \in \otimes_{j=1}^d \mathbb{R}^{n_j}$ it holds that

$$(1.17) \qquad\qquad\qquad \|A\|_F^2 \leq \|A\|_2 \|A\|_*.$$

Our next result characterizes tensors achieving equality in (1.17).

THEOREM 1.13. *The following two properties are equivalent for a nonzero tensor* A *in* $\otimes_{j=1}^d \mathbb{R}^{n_j}$ *or* $\mathrm{Sym}^d(\mathbb{R}^n)$:
   (i) A *is critical,*
   (ii) $\|A\|_2 \|A\|_* = \|A\|_F^2$.

We remark that the fact that extremal tensors achieve equality in (1.17) has been already proven in [8, Theorems 2.2 and 3.1].

1.4. Decomposition of Chebyshev forms. For symmetric tensors, the statement of Theorem 1.11 can be reinterpreted in terms of homogeneous forms. Note that a symmetric rank-one tensor $Y = \lambda\, y \otimes \cdots \otimes y$, $\lambda \in \mathbb{R}$, $\|y\| = 1$, is a symmetric best rank-one approximation to the symmetric tensor associated to a homogeneous form p if and only if

$$(1.18) \qquad \lambda = p(y) = \pm \|p\|_\infty.$$

Also, by (2.10), the homogeneous form associated to such a rank-one tensor is proportional to the dth power of a linear form,

$$p_Y(x) = \lambda \langle y, x \rangle^d = \lambda (y_1 x_1 + \cdots + y_n x_n)^d.$$

Therefore, in analogy to (1.13), Theorem 1.11 states that a form $p \in P_{d,n}$ is critical for the ratio $\|p\|_\infty / \|p\|_B$ if and only if it can be written as

$$(1.19) \qquad \left(\frac{\|p\|_\infty}{\|p\|_B} \right)^2 p(x) = \sum_{\ell=1}^{r} \alpha_\ell \lambda_\ell \langle y^\ell, x \rangle^d, \quad \sum_{\ell=1}^{r} \alpha_\ell = 1,\ \alpha_1, \ldots, \alpha_r > 0,$$

where $\lambda_i \in \mathbb{R}$ and $y^i \in \mathbb{R}^n$, $\|y^i\| = 1$, satisfy (1.18) for $i = 1, \ldots, r$.

From Theorem 1.1 we know that the binary Chebyshev forms $\mathrm{Ч}_{d,2}$ are extremal in $P_{2,d}$ and therefore they must admit a decomposition like (1.19). In Theorem 1.14 we provide such a decomposition. For $k = 0, \ldots, d-1$ denote $\theta_k = \pi k / d$ and $a_k = \cos(\theta_k)$, $b_k = \sin(\theta_k)$. Then $a_k + i b_k = e^{i\theta_k}$, $k = 0, \ldots, d-1$, are $2d$th roots of unity.

THEOREM 1.14. *For any $d \geq 1$ we have*

$$(1.20) \qquad \frac{1}{2^{d-1}} \mathrm{Ч}_{d,2}(x_1, x_2) = \frac{1}{d} \sum_{k=0}^{d-1} (-1)^k \left(x_1 a_k + x_2 b_k \right)^d$$

or, in polar coordinates,

$$(1.21) \qquad \frac{1}{2^{d-1}} \mathrm{Ч}_{d,2}(\cos\theta, \sin\theta) = \frac{1}{2^{d-1}} \cos(d\theta) = \frac{1}{d} \sum_{k=0}^{d-1} (-1)^k \cos(\theta - \theta_k)^d.$$

The second equality in (1.21) constitutes an interesting trigonometric identity, which we were not able to find in the literature.

In the following corollary of Theorem 1.14 we provide a decomposition (1.19) for cubic Chebyshev forms $\mathrm{Ч}_{3,n}$, which shows that they are critical in $P_{3,n}$.

COROLLARY 1.15. *For $n \geq 2$ we have*

(1.22)

$$\frac{1}{3n-2} \mathrm{Ч}_{3,n}(x) = \left(\frac{n+2}{9n-6} \right) x_1^3 + \frac{4}{9n-6} \sum_{i=2}^{n} - \left(\frac{x_1 + \sqrt{3} x_i}{2} \right)^3 + \left(\frac{-x_1 + \sqrt{3} x_i}{2} \right)^3.$$

In particular, $\mathrm{Ч}_{3,n}$, $n \geq 2$, is critical for the ratio $\|p\|_\infty / \|p\|_B$, $p \in P_{3,n}$.

In section 3.6 we use Corollary 1.15 to prove Theorem 1.3, that is, that $\mathrm{Ч}_{3,n}$ is a local minimum for the norm ratio.

It is interesting to note that a decomposition of $Ч_{3,n}$ or, more precisely, of its representing symmetric tensor, into nonsymmetric best rank-one approximations is trivially obtained. By (1.3), the associated symmetric tensor is

$$(1.23) \qquad A_n = e_1 \otimes e_1 \otimes e_1 - \sum_{k=2}^{n} (e_1 \otimes e_k \otimes e_k + e_k \otimes e_1 \otimes e_k + e_k \otimes e_k \otimes e_1),$$

where $e_1, \ldots, e_n$ denote the basic unit vectors in $\mathbb{R}^n$. Since $\|A_n\|_2 = 1$ by (1.5), this "decomposition into entries" is a decomposition into best rank-one approximations with equal weights. Scaling by $\|A_n\|_2^2 / \|A_n\|_F^2 = 1/(3n-2)$ provides a desired convex decomposition (1.13). While this proves that A_n is critical in $\mathbb{R}^n \otimes \mathbb{R}^n \otimes \mathbb{R}^n$ (see Theorem 1.10), it does not imply by itself that A_n is critical in $\mathrm{Sym}^3(\mathbb{R}^n)$. Thus Corollary 1.15 is a stronger statement. Observe also that (1.23) is a decomposition into pairwise orthogonal rank-one tensors. This together with (1.5) and (1.11) shows that the tensor A_n associated with the cubic Chebyshev form $Ч_{3,n}$ has orthogonal rank $3n - 2$.

2. Preliminaries. In this section we gather some basic definitions and preliminary results upon which we base our arguments for proving the main results in section 3.

2.1. Tensors, forms, and their norms. The space of $(n_1, \ldots, n_d)$-tensors is isomorphic to the space of multilinear maps on $\mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_d}$. The map associated to a tensor A is given by

$$(2.1) \qquad \left(x^{(1)}, \ldots, x^{(d)} \right) \mapsto \left\langle A, x^{(1)} \otimes \cdots \otimes x^{(d)} \right\rangle_F = \sum_{i_1, \ldots, i_d = 1}^{n_1, \ldots, n_d} a_{i_1 \ldots i_d} x_{i_1}^{(1)} \ldots x_{i_d}^{(d)}.$$

The spectral norm (1.7) of A equals the uniform norm of the restriction of the associated multilinear map to the product of unit spheres in $\mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_d}$.

As for the nuclear norm defined in (1.16), it can be shown that the infimum is always attained (see [10, Proposition 3.1]), and a decomposition $A = \sum_{\ell=1}^{r} X_\ell$ of A into rank-one tensors such that $\|A\|_* = \sum_{\ell=1}^{r} \|X_\ell\|_F$ is called *a nuclear decomposition*. We have already stated that the spectral and the nuclear norms are dual to each other, that is,

$$(2.2) \qquad \|A\|_2 = \max_{\|A'\|_* \le 1} |\langle A, A' \rangle_F|, \quad \|A\|_* = \max_{\|A'\|_2 \le 1} |\langle A, A' \rangle_F|,$$

and the three above introduced norms satisfy

$$(2.3) \qquad \|A\|_2 \le \|A\|_F, \quad \|A\|_F \le \|A\|_*, \quad \text{and} \quad \|A\|_F^2 \le \|A\|_2 \|A\|_*.$$

Moreover, in the first two inequalities in (2.3) equality holds if and only if A is a rank-one tensor. We refer to [6, 10] for these statements.

The product of orthogonal groups $O(n_1, \ldots, n_d) = O(n_1) \times \cdots \times O(n_d)$, whose elements are denoted $(\rho^{(1)}, \ldots, \rho^{(d)})$, acts on the space $\otimes_{j=1}^{d} \mathbb{R}^{n_j}$ as

$$(2.4) \qquad \left(\rho^{(1)}, \ldots, \rho^{(d)} \right) \cdot A = \left(\sum_{j_1, \ldots, j_d = 1}^{n_1, \ldots, n_d} \rho_{i_1 j_1}^{(1)} \ldots \rho_{i_d j_d}^{(d)} a_{j_1 \ldots j_d} \right),$$

preserving the Frobenius inner product and the spectral and the nuclear norms.

The $\binom{n+d-1}{d}$-dimensional space $\mathrm{Sym}^d(\mathbb{R}^n) \subset \otimes_{j=1}^d \mathbb{R}^n$ of symmetric n^d-tensors is isomorphic to the space $P_{d,n}$ of n-ary d-homogeneous real forms. The symmetric tensor A is identified with the form p_A defined as

$$(2.5) \qquad p_A(x) = \langle A, x \otimes \cdots \otimes x \rangle_F = \sum_{i_1,\ldots,i_d=1}^n a_{i_1 \ldots i_d} x_{i_1} \ldots x_{i_d}, \quad x \in \mathbb{R}^n,$$

which equals the restriction of the multilinear map (2.1) to the "diagonal" in $\mathbb{R}^n \times \cdots \times \mathbb{R}^n$. It is convenient to represent p_A in the basis of monomials

$$p_A(x) = \sum_{|\alpha|=d} a_\alpha x^\alpha,$$

where

$$(2.6) \qquad a_\alpha = \binom{d}{\alpha} a_{i_1 \ldots i_d}$$

and $\{i_1, \ldots, i_d\}$ is any collection of indices such that for $i = 1, \ldots, n$ the value i occurs α_i times among $i_1, \ldots, i_d$.

As an example, the binary Chebyshev form $Ч_{d,2}$ in (0.1) corresponds to the symmetric tensor with entries

$$(2.7) \qquad a_{i_1 \ldots i_d} = \begin{cases} (-1)^k & \text{if } \#\{i_j = 2\} = 2k, \\ 0 & \text{otherwise} \end{cases}$$

and the associated multilinear map (2.1) is given by

$$\left\langle A, x^{(1)} \otimes \cdots \otimes x^{(d)} \right\rangle = \sum_{k=0}^{[d/2]} (-1)^k \sum_{\#\{i_j=2\}=2k} x_{i_1}^{(1)} \ldots x_{i_d}^{(d)}.$$

Banach proved [1] that for a symmetric coefficient tensor A, the maximum absolute value of the multilinear form (2.1) on a product of spheres can be attained at diagonal inputs, in other words,

$$(2.8) \qquad \|A\|_2 = \max_{\|x\|=1} |p_A(x)| = \|p_A\|_\infty.$$

This is a generalization of the fact that for a symmetric matrix A the maximum absolute value of the bilinear form $x^T A y$ is, modulo scaling, attained when $x = y$ is an eigenvector for the eigenvalue with the largest absolute value. Therefore, spectral norm for symmetric tensors may be intrinsically defined in the space $\mathrm{Sym}^d(\mathbb{R}^n)$.

Next, one can easily check that the Frobenius inner product between two symmetric tensors $A = (a_{i_1 \ldots i_d})$, $A' = (a'_{i_1 \ldots i_d}) \in \mathrm{Sym}^d(\mathbb{R}^n)$ equals the *Bombieri product* between the corresponding homogeneous forms $p_A(x) = \sum_{|\alpha|=d} a_\alpha x^\alpha$ and $p_{A'}(x) = \sum_{|\alpha|=d} a'_\alpha x^\alpha$ with coefficients defined through (2.6):

$$(2.9) \qquad \langle A, A' \rangle_F = \sum_{i_1,\ldots,i_d=1}^n a_{i_1 \ldots i_d} a'_{i_1 \ldots i_d} = \sum_{|\alpha|=d} \binom{d}{\alpha}^{-1} a_\alpha a'_\alpha =: \langle p_A, p_{A'} \rangle_B.$$

By (2.8) and (2.9), the isomorphism $A \mapsto p_A$ establishes an isometry between $(\mathrm{Sym}^d(\mathbb{R}^n), \|\cdot\|_2)$ and $(P_{d,n}, \|\cdot\|_\infty)$, as well as between $(\mathrm{Sym}^d(\mathbb{R}^n), \|\cdot\|_F)$ and $(P_{d,n}, \|\cdot\|_B)$.

When $n_1 = \cdots = n_d = n$ the diagonal subaction of the action (2.4) preserves the subspace $\mathrm{Sym}^d(\mathbb{R}^n)$ of symmetric tensors and it corresponds to the action of the orthogonal group on the space $P_{d,n}$ of homogeneous forms by orthogonal change of variables:

$$\rho \in O(n), \ p \in P_{d,n} \ \mapsto \ \rho^* p \in P_{d,n}, \quad (\rho^* p)(x) = p(\rho^{-1} x).$$

Due to (2.9), this shows that the Bombieri inner product is invariant under such a change of variables.

Finally, we have already noted that according to (2.5) a symmetric rank-one tensor $Y = \pm y \otimes \cdots \otimes y$ corresponds to the dth power of a linear form $\langle y, \cdot \rangle$ as follows:

$$(2.10) \qquad p_Y(x) = \langle \pm y \otimes \cdots \otimes y, x \otimes \cdots \otimes x \rangle_F = \pm \langle y, x \rangle^d.$$

Hence a decomposition of a symmetric tensor into symmetric rank-one tensors corresponds to a decomposition of the associated homogeneous form into powers of linear forms. Note that by (2.5) the Bombieri inner product of any homogeneous form $p \in P_{d,n}$ with a dth power of a linear form $\langle y, \cdot \rangle$ equals $\langle p, \langle y, \cdot \rangle^d \rangle_B = p(y)$.

2.2. Best rank-one approximation ratio. Given a nonzero tensor $A \in \otimes_{j=1}^d \mathbb{R}^{n_j}$, a rank-one $(n_1, \ldots, n_d)$-tensor $Y = \lambda\, y^{(1)} \otimes \cdots \otimes y^{(d)}$, where $\lambda \in \mathbb{R}$ and $\|y^{(i)}\| = 1$, $i = 1, \ldots, d$, is a best rank-one approximation to A if and only if

$$(2.11) \qquad \lambda = \left\langle A, y^{(1)} \otimes \cdots \otimes y^{(d)} \right\rangle_F = \pm \|A\|_2.$$

Banach's result [1] implies that one can take $y^{(1)} = \cdots = y^{(d)} \in \mathbb{R}^n$ in (2.11) if the tensor $A \in \mathrm{Sym}^d(\mathbb{R}^n)$ is symmetric.

Also if $Y = \lambda\, y^{(1)} \otimes \cdots \otimes y^{(d)}$ is a best rank-one approximation of A as above, then for every $j = 1, \ldots, d$ the linear form $x^{(j)} \mapsto \langle A, y^{(1)} \otimes \cdots \otimes x^{(j)} \otimes \cdots \otimes y^{(d)} \rangle_F$ constrained to $\|x^{(j)}\| = 1$ achieves its maximum at $y^{(j)}$ and hence it vanishes on the orthogonal complement of $y^{(j)}$, that is,

$$(2.12) \qquad \left\langle A, y^{(1)} \otimes \cdots \otimes y^{(j-1)} \otimes x^{(j)} \otimes y^{(j+1)} \otimes \cdots \otimes y^{(d)} \right\rangle_F = 0$$

for all $x^{(j)} \in \mathbb{R}^{n_j}$ that are orthogonal to $y^{(j)}$.

We continue with some remarks on extremal tensors and best rank-one approximation ratio. From the definition (1.8) of a best rank-one approximation and (2.11) we have

$$\min_{\mathrm{rank}(X)=1} \|A - X\|_F^2 = \|A - Y\|_F^2 = \|A\|_F^2 - \|A\|_2^2$$

for any best rank-one approximation Y to $A \in \otimes_{j=1}^d \mathbb{R}^{n_j}$. Recalling the definition (1.9) of the best rank-one approximation ratio $\mathcal{A}(\otimes_{j=1}^d \mathbb{R}^{n_j})$, the maximum relative distance of a tensor to the set of rank-one tensors is given as

$$(2.13) \qquad \max_{0 \neq A \in \otimes_{j=1}^d \mathbb{R}^{n_j}} \min_{\mathrm{rank}(X)=1} \frac{\|A - X\|_F}{\|A\|_F} = \sqrt{1 - \mathcal{A}\left(\otimes_{j=1}^d \mathbb{R}^{n_j}\right)^2}$$

and is achieved for extremal tensors. This relation explains the name "best rank-one approximation ratio" for the constant $\mathcal{A}(\otimes_{j=1}^d \mathbb{R}^{n_j})$. When restricting to symmetric tensors, (2.13) holds with $\mathcal{A}(\mathrm{Sym}^d(\mathbb{R}^n))$ instead.

For context we note that the relation (2.13) shows that lower bounds on $\mathcal{A}(\otimes_{j=1}^d \mathbb{R}^{n_j})$ can be used for convergence analysis of greedy methods for low-rank approximation using rank-one tensors as a dictionary. For example, the pure greedy method to approximate $A \in \otimes_{j=1}^d \mathbb{R}^{n_j}$ produces a recursive sequence $A_{\ell+1} = A_\ell + Y_\ell$, where $A_0 = 0$ and Y_ℓ is a best rank-one approximation of $A - A_\ell$. Then (2.13) implies

$$\|A - A_{\ell+1}\|_F \leq \sqrt{1 - \mathcal{A}\left(\otimes_{j=1}^d \mathbb{R}^{n_j}\right)^2}\|A - A_\ell\|_F;$$

see [19] for a general introduction to greedy methods. For a more general problem of finding an approximate low-rank minimizer for a smooth cost function $f\colon \otimes_{j=1}^d \mathbb{R}^{n_j} \to \mathbb{R}$, one could replace Y_ℓ with a (scaled) best rank-one approximation of a suitable residual, for example, the negative gradient $-\nabla f(A_\ell)$. Then $\mathcal{A}(\otimes_{j=1}^d \mathbb{R}^{n_j})$ is a lower bound for the (cosine of the) angle between the search direction Y_ℓ and $-\nabla f(A_\ell)$ and hence can be used to estimate the convergence of such an iteration; see, e.g., [20] and references therein. Again, for symmetric tensors once can replace $\mathcal{A}(\otimes_{j=1}^d \mathbb{R}^{n_j})$ with $\mathcal{A}(\mathrm{Sym}^d(\mathbb{R}^n))$ in these considerations.

In the following lemma we show that the best rank-one approximation ratio strictly decreases with the dimension.

LEMMA 2.1. *Let $\mathcal{A}_{d,n}$ denote either $\mathcal{A}(\otimes_{j=1}^d \mathbb{R}^n)$ or $\mathcal{A}(\mathrm{Sym}^d(\mathbb{R}^n))$. Then for any $d \geq 1$ and $n \geq 1$ we have*

$$\mathcal{A}_{d,n+1} \leq \frac{\mathcal{A}_{d,n}}{\sqrt{1 + \mathcal{A}_{d,n}^2}}.$$

Proof. Let $A \in \otimes_{j=1}^d \mathbb{R}^n$ be an n^d-tensor of unit Frobenius norm, $\|A\|_F = 1$. For $\varepsilon \in [0, 1]$, let $A^\varepsilon \in \otimes_{j=1}^d \mathbb{R}^{n+1}$ be the $(n+1)^d$-tensor with entries

$$a_{i_1 \ldots i_d}^\varepsilon = \begin{cases} \sqrt{1 - \varepsilon^2 \|A\|_2^2}\, a_{i_1 \ldots i_d} & \text{if } i_1, \ldots, i_d \leq n, \\ \varepsilon \|A\|_2 & \text{if } i_1 = \cdots = i_d = n+1, \\ 0 & \text{otherwise.} \end{cases}$$

Observe that $\|A^\varepsilon\|_F = 1$, and A^ε is symmetric if A is. Let $\xi^{(1)}, \ldots, \xi^{(d)}$ be unit norm vectors in $\mathbb{R}^{n+1}$ partitioned as $\xi^{(j)} = (x^{(j)}, z^{(j)})$ with $x^{(j)} \in \mathbb{R}^n$ and $z^{(j)} \in \mathbb{R}$. Then from the "block diagonal" structure of A^ε it follows that

$$\langle A^\varepsilon, \xi^{(1)} \otimes \cdots \otimes \xi^{(d)} \rangle_F = \sqrt{1 - \varepsilon^2 \|A\|_2^2}\langle A, x^{(1)} \otimes \cdots \otimes x^{(d)} \rangle_F + \varepsilon \|A\|_2 z^{(1)} \cdots z^{(d)}$$

$$\leq \max\left(\sqrt{1 - \varepsilon^2 \|A\|_2^2}, \varepsilon\right) \|A\|_2 (\|x^{(1)}\| \cdots \|x^{(d)}\| + z^{(1)} \cdots z^{(d)}).$$

By a generalized Hölder inequality [12, section 11], the term in the right brackets is bounded by one. The maximum on the left, on the other hand, takes its minimal value for $\varepsilon = 1/\sqrt{1 + \|A\|_2^2}$. Since $\xi^{(1)}, \ldots, \xi^{(d)}$ were arbitrary, this shows

$$\|A^\varepsilon\|_2 \leq \frac{\|A\|_2}{\sqrt{1 + \|A\|_2^2}}.$$

The assertions follow by choosing A to be an extremal tensor in the space $\otimes_{j=1}^d \mathbb{R}^n$ or $\mathrm{Sym}^d(\mathbb{R}^n)$, respectively. $\qquad\square$

The previous lemma provides a lower bound on the rank of extremal tensors. Recall that the *(real) rank* of a tensor $A \in \otimes_{j=1}^{d} \mathbb{R}^{n_j}$ is the smallest number r that is needed to represent A as the linear combination

(2.14) $$A = X_1 + \cdots + X_r$$

of rank-one tensors $X_1, \ldots, X_r$. The *(real) symmetric rank* of a symmetric tensor A is the smallest number of symmetric rank-one tensors needed for (2.14) to hold.

PROPOSITION 2.2. *If $A \in \otimes_{j=1}^{d} \mathbb{R}^n$ is an extremal tensor, its rank must be at least n. If $A \in \mathrm{Sym}^d(\mathbb{R}^n)$ is an extremal symmetric tensor, its symmetric rank must be at least n.*

Proof. Let $A \in \otimes_{j=1}^{d} \mathbb{R}^n$ be a tensor of rank at most $n-1$, that is,

$$A = v_1^{(1)} \otimes \cdots \otimes v_1^{(d)} + \cdots + v_{n-1}^{(1)} \otimes \cdots \otimes v_{n-1}^{(d)}.$$

For $j = 1, \ldots, d$ let $V^{(j)} \simeq \mathbb{R}^{n-1}$ be any $(n-1)$-dimensional subspace of $\mathbb{R}^n$ that contains vectors $v_1^{(j)}, \ldots, v_{n-1}^{(j)}$. Since $A \in V^{(1)} \otimes \cdots \otimes V^{(d)} \simeq \otimes_{j=1}^{d} \mathbb{R}^{n-1}$ we have

$$\frac{\|A\|_2}{\|A\|_F} \geq \mathcal{A}\left(\otimes_{j=1}^{d} \mathbb{R}^{n-1}\right).$$

Thus, by Lemma 2.1, A cannot be extremal in $\otimes_{j=1}^{d} \mathbb{R}^n$.

When A is symmetric and of symmetric rank at most $n-1$ we can choose $V^{(1)} = \cdots = V^{(d)} = V$ so that $A \in \mathrm{Sym}^d(V) \simeq \mathrm{Sym}^d(\mathbb{R}^{n-1})$, leading to the analogous conclusion. $\quad\square$

2.3. Generalized gradients and local optimality of Lipschitz functions. The problem of determining the best rank-one approximation ratio of a tensor space and finding extremal tensors is a constrained optimization problem for a Lipschitz function. The theory of generalized gradients developed by Clarke [4] provides necessary optimality conditions. We provide here only the most necessary facts of this theory needed for our results. A comprehensive introduction is given, e.g., in [5].

A function $f : \mathbb{R}^m \to \mathbb{R}$ is called *Lipschitz*, if there exist a constant L such that $|f(p) - f(q)| \leq L\|p - q\|$ for all pairs $p, q \in \mathbb{R}^m$. By the classical Rademacher's theorem, a Lipschitz function f is differentiable at almost all (in the sense of Lebesgue measure) points $p \in \mathbb{R}^m$. Denote by $\nabla f(p)$ the gradient of f at such a point. The *generalized gradient* of f at any $p \in \mathbb{R}^m$, denoted as $\partial f(p)$, is then defined as the convex hull of the set of all limits $\nabla f(p_i)$, where p_i is a sequence of differentiable points that converges to p. It turns out that $\partial f(p)$ is a nonempty convex compact subset of $\mathbb{R}^m$. Moreover $\partial f(p)$ is a singleton if and only if f is differentiable at p, in which case $\partial f(p) = \{\nabla f(p)\}$.

Let S be a differentiable submanifold in $\mathbb{R}^m$. Then a necessary condition for the Lipschitz function f to attain a local minimum relative to S at $x \in S$ is that

(2.15) $$\partial f(p) \cap N_S(p) \neq \emptyset,$$

where $N_S(p)$ denotes the normal space, that is, the orthogonal complement of the tangent space of S at p. Note that this is a "Lipschitz" analogue of the classical Lagrange multipliers rule for continuously differentiable functions. We refer to [5, section 2.4]. Every point $p \in S$ that satisfies (2.15) is called a *critical point* of f on S. Hence local minima of f on S are among the critical points.

The proofs of Theorems 1.10 and 1.11 in section 3.4 consist in applying the necessary optimality condition (2.15) to the spectral norm function on the sphere defined by Frobenius norm. Here two things are of relevance. First, for a Euclidean sphere S we have $N_s(p) = \{\mu p : \mu \in \mathbb{R}\}$. Hence the condition (2.15) becomes

$$\mu p \in \partial f(p) \tag{2.16}$$

for some $\mu \in \mathbb{R}$. Second, by (1.7), the spectral norm is an example of a so-called max function, that is, a function of the type

$$f(p) = \max_{u \in C} g(p, u), \tag{2.17}$$

where C is compact. Under certain smoothness conditions on the function g, which are satisfied for spectral norm (1.7), Clarke [4, Theorem 2.1] has determined the following characterization of the generalized gradient:

$$\partial f(p) = \mathrm{conv}\{\nabla_p g(p, u) : u \in M(p)\}, \tag{2.18}$$

where conv denotes the convex hull and $M(p)$ is the set of all maximizers u in (2.17) for a fixed p. For the spectral norm (1.7), this set consists of all normalized best rank-one approximations of a given tensor; see (3.4).

3. Proof of main results. Our main results are proved in this section. We are going to repeatedly use the equivalence (2.5) between symmetric tensors and homogeneous forms and the corresponding relations (2.8), (2.9) for the different norms.

3.1. Binary forms. This subsection is devoted to the proof of Theorem 1.1. While the given proof is self-contained, some arguments could be omitted with reference to results in [15].

Proof of Theorem 1.1. By (1.4),

$$\frac{\|Ч_{d,2}\|_2}{\|Ч_{d,2}\|_B} = \frac{1}{\sqrt{2^{d-1}}}.$$

It then follows from (1.12) that this value equals $\mathcal{A}(\otimes_{j=1}^d \mathbb{R}^2)$, so the symmetric tensor associated to the Chebyshev form must be extremal both in $\otimes_{j=1}^d \mathbb{R}^2$ and in $\mathrm{Sym}^d(\mathbb{R}^2)$.

We now consider the uniqueness statements. When $d = 1$, the space $P_{1,n}$ consists of linear forms $p(x) = \langle a, x \rangle$, for any of which it holds that $\|p\|_\infty/\|p\|_B = 1$. In the case $d = 2$ of quadratic forms, the minimal ratio between spectral and Frobenius norm of a symmetric $n \times n$ matrix is attained for multiples of symmetric orthogonal matrices only and takes the value $1/\sqrt{n}$. When $n = 2$, all such matrices can be obtained by orthogonal transformation and scaling from the two diagonal matrices with diagonal entries $(1, 1)$ and $(1, -1)$, respectively. This corresponds to the asserted quadratic forms $p \in P_{d,2}$.

In the case $d \geq 3$ we have to show that the only symmetric 2^d-tensors A satisfying

$$\|A\|_2 = 1, \quad \|A\|_F = \sqrt{2^{d-1}} \tag{3.1}$$

are obtained from orthogonal transformations of the Chebyshev form $Ч_{d,2}$. To this end, we show that under the additional condition

$$p_A(e_1) = \langle A, e_1 \otimes \cdots \otimes e_1 \rangle_F = 1 = \|A\|_2, \tag{3.2}$$

the form p_A equals $Ч_{d,2}$. The proof is given by induction over $d \geq 3$. Before giving this proof we note that for a 2^d-tensor A satisfying (3.1), its two *slices* $A_1 = (a_{i_1 \ldots i_{d-2} 1})$ and $A_2 = (a_{i_1 \ldots i_{d-1} 2})$ necessarily have the same Frobenius norm $\|A_1\|_F = \|A_2\|_F = \sqrt{2^{d-2}}$. In fact, $\|A\|_2 = 1$ implies $\|A_1\|_2 \leq 1$ and hence, by (1.12), $\|A_1\|_F \leq \sqrt{2^{d-2}}$. Since the same holds for A_2 and since $\|A\|_F^2 = \|A_1\|_F^2 + \|A_2\|_F^2$, the claim follows. Moreover, $\|A_1\|_2 = \|A_2\|_2 = 1$, again by (1.12), so that both slices are necessarily extremal. Note that by the same argument, every $2^{d'}$-subtensor of A with $d' < d$ must be extremal.

We begin the induction with $d = 3$. Assume $A \in \mathrm{Sym}^3(\mathbb{R}^2)$ satisfies (3.1) and (3.2). Then we have seen that both, say, frontal slices of A are themselves extremal symmetric 2×2 matrices. By (3.2), $a_{111} = 1$ and the tensor $e_1 \otimes e_1 \otimes e_1$ is a best rank-one approximation. From (2.12) we then deduce that $a_{112} = a_{121} = a_{211} = 0$. The only two remaining options for the slices of A are

$$A = \begin{pmatrix} 1 & 0 & | & 0 & \pm 1 \\ 0 & \pm 1 & | & \pm 1 & 0 \end{pmatrix}.$$

But the case $a_{122} = a_{221} = a_{212} = +1$ is also not possible, since it corresponds to the form $p_A(x) = x_1^3 + 3x_1 x_2^2$ whose maximum on the sphere is $\|p_A\|_\infty = \sqrt{2} > 1$. Therefore, $a_{122} = a_{221} = a_{212} = -1$ and $p_A = x_1^3 - 3x_1 x_2^2$ is the cubic Chebyshev form.

We proceed with the induction step. If $A \in \mathrm{Sym}^{d+1}(\mathbb{R}^2)$ satisfies (3.1) and (3.2), then its two slices $A_1 = (a_{i_1 \ldots i_d 1})$ and $A_2 = (a_{i_1 \ldots i_d 2})$ are extremal 2^d-tensors. Since $p_{A_1}(e_1) = p_A(e_1) = 1$, it follows from the induction hypothesis that $A_1 = Ч_{d,2}$. So its entries are given by (2.7). Let $a_{i_1 \ldots i_d 2}$ denote an entry of the second slice. Due to the symmetry of A, every entry in the second slice, except for the entry $a_{2 \ldots 2}$, equals an entry in the first slice after a permutation of the indices. Since this permutation does not affect the number of occurrences of the value 2 among the indices, the definition (2.7) applies to all these entries as well. It remains to show that the entry $a_{2 \ldots 2}$ satisfies (2.7), that is, equals zero in case $d + 1$ is odd, and equals $(-1)^m$ in case $d + 1 = 2m$ is even. This entry is part of the symmetric subtensor $A' = (a_{i_1 i_2 i_3 2 \ldots 2})$, which as we have noted above must be extremal as well. Since the entries of the first slice A_1 are given by (2.7), we find that

$$p_{A'}(x) = (-1)^{m-1} \left(x_1^3 - 3x_1 x_2^2 \right) + a_{2 \ldots 2} x_2^3$$

if $d + 1 = 2m + 1$ is odd. Since A' is extremal, it then follows from the base case $d = 3$ that $a_{2 \ldots 2} = 0$. In case $d + 1 = 2m$ is even, we get

$$p_{A'}(x_1, x_2) = (-1)^{m-1} 3x_1^2 x_2 + a_{2 \ldots 2} x_2^3,$$

which by a small consideration implies $a_{2 \ldots 2} = (-1)^m$. This concludes the proof. $\quad\square$

3.2. Ternary cubic tensors. In this section we prove Theorem 1.5. It has been mentioned in section 1.2 how Corollary 1.6 follows from it, and that the statement of Theorem 1.2 is included in the latter.

The proof of Theorem 1.5 requires a fact from [13]. Since it is not explicitly formulated there, we state it here as a lemma.

LEMMA 3.1. *For odd n let $A_1, A_2 \in \mathbb{R}^n \otimes \mathbb{R}^n$ be two $n \times n$ matrices. If at least one of them is invertible, then there exist orthogonal matrices $\rho, \rho' \in O(n)$ such that*

$$\rho A_1 \rho' = \begin{pmatrix} B_1 & c_1 \\ 0 & d_1 \end{pmatrix}, \quad \rho A_2 \rho' = \begin{pmatrix} B_2 & c_2 \\ 0 & d_2 \end{pmatrix},$$

where B_1, B_2 are matrices of size $(n-1) \times (n-1)$, c_1, c_2 are $(n-1)$-dimensional vectors and d_1, d_2 are real numbers.

Proof. We can assume A_1 is invertible. Since n is odd, the matrix $A_1^{-1}A_2$ has at least one real eigenvalue d. Then there exists an invertible matrix P such that

$$P^{-1}A_1^{-1}A_2P = \begin{pmatrix} B & c \\ 0 & d \end{pmatrix},$$

where B is a matrix of size $(n-1) \times (n-1)$ and c is an $(n-1)$-dimensional vector. Consider QR decompositions of A_1P and P, that is,

$$A_1P = Q_1R_1, \quad P = Q_2R_2,$$

where Q_1, Q_2 are orthogonal, and R_1, R_2 are upper triangular and invertible. We set $\rho = Q_1^{-1}$ and $\rho' = Q_2$. Then

$$\rho A_1 \rho' = R_1 P^{-1} A_1^{-1} A_1 P R_2^{-1} = R_1 R_2^{-1}$$

is the product of two upper block triangular matrices, hence upper block triangular. Similarly,

$$\rho A_2 \rho' = R_1 P^{-1} A_1^{-1} A_2 P R_2^{-1} = R_1 \begin{pmatrix} B & c \\ 0 & d \end{pmatrix} R_2^{-1}$$

has the asserted upper block triangular structure. $\qquad\square$

In [13] the previous lemma is used to show that for odd n the maximum possible orthogonal rank of an $(n, n, 2)$-tensor is $2n-1$. We will only need that the orthogonal rank of a $(3, 3, 2)$-tensor is not larger than 5, which actually follows quite easily from the lemma by applying it to the slices.

Proof of Theorem 1.5. Note that the aforementioned result of [14] that $1/\sqrt{7}$ is an upper bound for $\mathcal{A}(\mathbb{R}^3 \otimes \mathbb{R}^3 \otimes \mathbb{R}^3)$ in combination with (1.11) implies that the maxmimal orthogonal rank cannot be less than seven. We will show that it is at most seven.

For $A \in \mathbb{R}^3 \otimes \mathbb{R}^3 \otimes \mathbb{R}^3$, it is convenient to write $A = (A_1|A_2|A_3)$, where A_1, A_2, A_3 are the 3×3 slices along the third dimension. If none of the matrices A_1, A_2, A_3 is invertible, each of them can be decomposed into a sum of two rank-one matrices that are orthogonal in the Frobenius inner product: $A_i = u_i^{(1)} \otimes u_i^{(2)} + v_i^{(1)} \otimes v_i^{(2)}$, $i = 1, 2, 3$. This leads to a decomposition of A into at most six pairwise orthogonal rank-one tensors:

$$A = \sum_{i=1}^{3} u_i^{(1)} \otimes u_i^{(2)} \otimes e_i + v_i^{(1)} \otimes v_i^{(2)} \otimes e_i.$$

Assume without loss of generality that the first slice A_1 is invertible. Lemma 3.1 together with the invariance of orthogonal rank under orthogonal transformations (2.4) allows us to assume that A has the form

$$
\begin{aligned}
A &= \begin{pmatrix} * & * & * & * & * & * & * & * & * \\ * & * & * & * & * & * & * & * & * \\ 0 & 0 & * & 0 & 0 & * & * & * & * \end{pmatrix} \\
&= \begin{pmatrix} * & * & * & * & * & * & * & * & * \\ * & * & * & * & * & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & * & 0 & 0 & * & * & * & * \end{pmatrix}.
\end{aligned}
$$

The first term is essentially a $(2,3,3)$-tensor, so its orthogonal rank is at most five by the result of [13]. In particular, it has a decomposition into at most five pairwise orthogonal rank-one tensors with zero bottom rows. Since the bottom row of the second term is a rank-two matrix the orthogonal rank of A is at most seven. $\square$

3.3. On symmetric orthogonal tensors. We prove Proposition 1.7 below. For the general definition of orthogonal tensors of arbitrary size we refer to [15]. For n^d-tensors we can use the recursive definition that $A \in \otimes_{j=1}^{d} \mathbb{R}^n$ is orthogonal if $A \times_j u$ is orthogonal for every $j = 1, \ldots, d$ and every unit norm vector $u \in \mathbb{R}^n$, where for $d = 2$ we agree to the standard definition of an orthogonal matrix. Here and in the proof below we use standard notation $A \times_j u = (\sum_{i_j=1}^{n} a_{i_1 \ldots i_j \ldots i_d} u_{i_j})$ for partial contraction of a tensor A with a vector u along mode j, resulting in a tensor of order $d-1$. Note that the above definition implies that every $n^{d'}$-subtensor, $d' < d$, of A is itself orthogonal.

Proof of Proposition 1.7 *and Corollary* 1.8. It has been shown in [15] that an n^d-tensor A is orthogonal if and only if it satisfies $\|A\|_2 = 1$ and $\|A\|_F = \sqrt{n^{d-1}}$, and such tensors only exist when $n = 1, 2, 4, 8$. Therefore, the statement that for $n = 2$ the only symmetric orthogonal tensors are the ones obtained from the Chebyshev form $Ч_{d,2}$ is hence equivalent to Theorem 1.1. Also, Corollary 1.8 is immediate from Proposition 1.7.

We thus only have to show that for $n = 4, 8$ an orthogonal n^d-tensor cannot be symmetric. We only consider the case $n = 4$; the arguments for $n = 8$ are analogous. Since $n^{d'}$-subtensors of an orthogonal tensor are necessarily orthogonal, it is enough to show that orthogonal $4 \times 4 \times 4$ tensors cannot be symmetric. Assume to the contrary that such a tensor A exists. Then $\|A\|_2 = 1$ and A admits a symmetric best rank-one approximation of Frobenius norm one. Since orthogonality and symmetry are preserved under the action of $O(4)$ we can assume that $e_1 \otimes e_1 \otimes e_1$ is the best rank-one approximation of A, that is, $a_{111} = \langle A, e_1 \otimes e_1 \otimes e_1 \rangle_F = \|A\|_2 = 1$. On the other hand, the first frontal slice $A \times_3 e_1$ must be a symmetric orthogonal matrix, so it is of the form

$$A \times_3 e_1 = \begin{pmatrix} 1 & 0 \\ 0 & B \end{pmatrix},$$

where B is a symmetric orthgonal 3×3 matrix. By applying further orthogonal transformation that fix the vector e_1, we can assume that B is a diagonal matrix with diagonal entries $\varepsilon_1, \varepsilon_2, \varepsilon_3 \in \{+1, -1\}$. Since A is symmetric and in fact every slice has to be an orthogonal matrix, we find that $A = (A \times_3 e_1 | A \times_3 e_2 | A \times_3 e_3 | A \times_3 e_4)$ is of the form

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & \varepsilon_1 & 0 & 0 & 0 & 0 & \varepsilon_3 & 0 & 0 & 0 & 0 & \varepsilon_4 \\ 0 & \varepsilon_1 & 0 & 0 & \varepsilon_1 & 0 & 0 & 0 & 0 & 0 & 0 & \varepsilon_0 & 0 & 0 & \varepsilon_0 & 0 \\ 0 & 0 & \varepsilon_2 & 0 & 0 & 0 & 0 & \varepsilon_0 & \varepsilon_3 & 0 & 0 & 0 & 0 & \varepsilon_0 & 0 & 0 \\ 0 & 0 & 0 & \varepsilon_3 & 0 & 0 & \varepsilon_0 & 0 & 0 & \varepsilon_0 & 0 & 0 & \varepsilon_4 & 0 & 0 & 0 \end{pmatrix},$$

where also $\varepsilon_0 \in \{+1, -1\}$. For $i = 2, 3, 4$ the matrices $A \times_3 (e_1 + e_i)/\sqrt{2}$ must be orthogonal as well, which yields $\varepsilon_0 = 1$ and $\varepsilon_2 = \varepsilon_3 = \varepsilon_4 = -1$. But then the matrix

$$A \times_3 \left(\frac{e_1 - e_2}{\sqrt{2}} \right) = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & -1 & -1 \\ 0 & 0 & -1 & -1 \end{pmatrix}$$

is not orthogonal, which contradicts the assumption that A is an orthogonal tensor. □

3.4. Variational characterization. In Theorems 1.10 and 1.11 we characterize critical tensors in $\otimes_{j=1}^d \mathbb{R}^{n_j}$ and $\mathrm{Sym}^d(\mathbb{R}^n)$ in terms of decompositions into best rank-one approximations. We now prove these results and then derive Corollary 1.12. Afterward we prove Theorem 1.13.

Proof of Theorems 1.10 *and* 1.11. From section 2.3, specifically (2.16), it follows that a nonzero tensor $A' \in \otimes_{j=1}^d \mathbb{R}^{n_j}$ is critical in the sense of Definition 1.9 if the tensor $A = A'/\|A'\|_F$ of Frobenius norm one satisfies

$$(3.3) \qquad \mu A \in \partial \|A\|_2$$

for some $\mu \in \mathbb{R}$. By (1.7), the spectral norm is a max function of the type (2.17) which is easily shown to satisfy the conditions of [4, Theorem 2.1]. Therefore, its generalized derivative is given by the formula (2.18), which in the case of the max function (1.7) reads

$$(3.4) \qquad \partial \|A\|_2 = \mathrm{conv}\left\{X : \|X\|_F = 1, \mathrm{rank}(X) = 1, \langle A, X\rangle_F = \|A\|_2\right\},$$

where conv denotes the convex hull. This lets us write (3.3) as

$$(3.5) \qquad \mu A = \sum_{\ell=1}^r \alpha_\ell X_\ell,$$

where $r > 0$ is a natural number,[1] $\alpha_1, \ldots, \alpha_r > 0$ are such that $\alpha_1 + \cdots + \alpha_r = 1$, and X_ℓ are rank-one tensors of unit Frobenius norm satisfying $\langle A, X_\ell\rangle_F = \|A\|_2$. By taking the Frobenius inner product with A itself in (3.5), we find that

$$\mu = \frac{\|A\|_2}{\|A\|_F^2}.$$

Therefore, after multiplying the resulting equation (3.5) by $\|A\|_2$ we obtain the asserted statement of Theorem 1.10, since, by (2.11), the rank-one tensors $Y_\ell = \|A\|_2 X_\ell$ are best rank-one approximations of A.

Considering symmetric tensors instead of general ones in the previous arguments yields a proof of Theorem 1.11. Here it is crucial that in the definition (1.7) of spectral norm for symmetric tensors one can restrict to take the maximum over symmetric rank-one tensors of unit Frobenius norm thanks to Banach's theorem; cf. (2.8). □

Proof of Corollary 1.12. By Proposition 2.2 any extremal tensor in $\otimes_{j=1}^d \mathbb{R}^n$ or $\mathrm{Sym}^d(\mathbb{R}^n)$ must be of rank (respectively, symmetric rank) at least n. In particular, there cannot be less than n best rank-one approximations in the expansions (1.14) and (1.15). □

Proof of Theorem 1.13. Let a tensor A (either in $\otimes_{j=1}^d \mathbb{R}^{n_j}$ or in $\mathrm{Sym}^d(\mathbb{R}^n)$) be critical, that is, by Theorem 1.10, respectively, Theorem 1.11,

$$(3.6) \qquad A = \left(\frac{\|A\|_F}{\|A\|_2}\right)^2 \sum_{\ell=1}^r \alpha_\ell Y_\ell$$

---

[1]By the classical Carathéodory theorem one can take $r \leq \dim(\otimes_{j=1}^d \mathbb{R}^{n_j}) + 1 = n_1 \cdots n_d + 1$.

for some (symmetric, if A is symmetric) best rank-one approximations $Y_1, \ldots, Y_r$ to A, and coefficients $\alpha_1, \ldots, \alpha_r > 0$ that sum up to one. Recall from section 2.1 that the nuclear norm is dual to the spectral norm. By (2.2), this in particular means there exists a tensor A^* satisfying $\|A^*\|_2 \leq 1$ and $\|A\|_* = \langle A, A^* \rangle_F$. Note that we then have $\langle X, A^* \rangle_F \leq \|X\|_F \|A^*\|_2 \leq \|X\|_F$ for every rank-one tensor X. Since $\|Y_\ell\|_F = \|A\|_2$, it hence follows from (3.6) that

$$\|A\|_* = \langle A, A^* \rangle_F = \left(\frac{\|A\|_F}{\|A\|_2} \right)^2 \sum_{\ell=1}^r \alpha_\ell \langle Y_\ell, A^* \rangle_F \leq \frac{\|A\|_F^2}{\|A\|_2},$$

which is the converse inequality to (1.17). This shows that (i) implies (ii).

Assume now that (ii) holds for a nonzero tensor A, that is, $\|A\|_2 \|A\|_* = \|A\|_F^2$. By the definition of the nuclear norm there exist $r \in \mathbb{N}$, positive numbers $\beta_1, \ldots, \beta_r > 0$, and rank-one tensors $X_1, \ldots, X_r$ of unit Frobenius norm such that

$$(3.7) \qquad A = \sum_{\ell=1}^r \beta_\ell X_\ell \quad \text{and} \quad \|A\|_* = \sum_{\ell=1}^r \beta_\ell.$$

If A is symmetric, $X_1, \ldots, X_r$ can be taken symmetric [10]. Taking the Frobenius inner product with A in the first of these equations gives

$$\|A\|_2 \|A\|_* = \langle A, A \rangle_F = \sum_{\ell=1}^r \beta_\ell \langle A, X_\ell \rangle_F.$$

Since $\langle A, X_\ell \rangle_F \leq \|A\|_2$ for $\ell = 1, \ldots, r$ and since $\beta_1, \ldots, \beta_r$ sum up to $\|A\|_*$, this equality can hold only if $\langle A, X_\ell \rangle_F = \|A\|_2$ for $\ell = 1, \ldots, r$. Since, by (2.11), the rank-one tensors $Y_\ell = \|A\|_2 X_\ell$, $\ell = 1, \ldots, r$, are then best rank-one approximations of A, we see that (3.7) is equivalent to (3.6), which by Theorems 1.10 and 1.11 means that A is critical. $\qquad \square$

Remark 3.2. Observe from the proof that decomposition (3.6) of a critical tensor into its best rank-one approximations is also its nuclear decomposition. Vice versa, any nuclear decomposition of a tensor A satisfying $\|A\|_2 \|A\|_* = \|A\|_F^2$ can be turned into a convex linear combination of best rank-one approximations of the rescaled tensor $\|A\|_2^2 / \|A\|_F^2 \, A$.

3.5. Decomposition of Chebyshev forms. In this section we give the proof of Proposition 1.14, which realizes the decomposition of critical tensors into symmetric best rank-one approximations, that is, corresponding powers of linear forms, for the Chebyshev forms $Ч_{d,2}$.

Proof of Proposition 1.14. Recall that for any $k = 0, \ldots, d-1$ we denote $\theta_k = \pi k / d$ and $a_k = \cos(\theta_k)$, $b_k = \sin(\theta_k)$. Let us observe that for any such k we can write

$$\cos(d\theta) = \operatorname{Re}\left((-1)^k e^{id(\theta - \theta_k)} \right) = (-1)^k \sum_{\ell=0}^{[d/2]} \binom{d}{2\ell} (-1)^\ell \cos(\theta - \theta_k)^{d-2\ell} \sin(\theta - \theta_k)^{2\ell}$$

and therefore

$$\cos(d\theta) = \frac{1}{d} \sum_{\ell=0}^{[d/2]} \binom{d}{2\ell} (-1)^\ell \sum_{k=0}^{d-1} (-1)^k \cos(\theta - \theta_k)^{d-2\ell} \sin(\theta - \theta_k)^{2\ell}.$$

Below we show that for any $\ell = 0, \ldots, [d/2]$ it holds that

$$(3.8) \qquad (-1)^\ell \sum_{k=0}^{d-1} (-1)^k \cos(\theta - \theta_k)^{d-2\ell} \sin(\theta - \theta_k)^{2\ell} = \sum_{k=0}^{d-1} (-1)^k \cos(\theta - \theta_k)^d.$$

This together with the identity $\sum_{\ell=0}^{[d/2]} \binom{d}{2\ell} = 2^{d-1}$ implies (1.21) (and hence also (1.20)).

To derive (3.8) we write

$$(-1)^\ell \sum_{k=0}^{d-1} (-1)^k \cos(\theta - \theta_k)^{d-2\ell} \sin(\theta - \theta_k)^{2\ell}$$

$$= \sum_{k=0}^{d-1} (-1)^k \cos(\theta - \theta_k)^{d-2\ell} \sum_{j=0}^{\ell} \binom{\ell}{j} \cos(\theta - \theta_k)^{2j} (-1)^{\ell-j}$$

$$= \sum_{j=0}^{\ell} \binom{\ell}{j} (-1)^{\ell-j} \sum_{k=0}^{d-1} (-1)^k \cos(\theta - \theta_k)^{d-2(\ell-j)}$$

and claim that for $j = 0, \ldots, \ell - 1$ the inner sum in the last formula is zero. In fact, we will show that for $s = 1, \ldots, [d/2]$

$$(3.9) \qquad \sum_{k=0}^{d-1} (-1)^k \cos(\theta - \theta_k)^{d-2s} = 0.$$

For this let us observe first that Chebyshev polynomials of the first kind $T_{d-2j}(\cos \theta) = \cos((d-2j)\theta)$, $j = 1, \ldots, [d/2]$, form a basis in the space spanned by univariate real polynomials of degrees $d-2, d-4, \ldots, d-2[d/2]$. As a consequence one can express $\cos(\theta - \theta_k)^{d-2s}$ in terms of $T_{d-2j}(\cos(\theta - \theta_k))$ for $j = s, \ldots, [d/2]$, and thus in order to prove (3.9), it is enough to show that for $s = 1, \ldots, [d/2]$ we have

$$\sum_{k=0}^{d-1} (-1)^k \cos((d-2s)(\theta - \theta_k)) = 0.$$

But this follows from the identity

$$\sum_{k=0}^{d-1} (-1)^k e^{i(d-2s)(\theta - \theta_k)} = e^{i(d-2s)\theta} \sum_{k=0}^{d-1} \left(e^{i2\pi s/d} \right)^k = 0,$$

hence the proof is complete   □

We now derive Corollary 1.15 which, in particular, implies that the cubic Chebyshev forms Ч$_{3,n}$ are critical for the ratio $\|p\|_\infty / \|p\|_B$, $p \in P_{3,n}$.

Proof of Corollary 1.15. From Proposition 1.14 we get

(3.10)

$$Ч_{3,2}(x_1, x_2) = x_1^3 - 3x_1 x_2^2 = \frac{4}{3} \left(x_1^3 - \left(\frac{x_1 - \sqrt{3}x_2}{2} \right)^3 + \left(\frac{-x_1 + \sqrt{3}x_2}{2} \right)^3 \right).$$

We then write

$$\text{Ч}_{3,n}(x) = x_1^3 - 3x_1(x_2^2 + \cdots + x_n^2) = -(n-2)x_1^3 + \sum_{i=2}^{n} \left(x_1^3 - 3x_1 x_i^2 \right)$$

$$= -(n-2)x_1^3 + \frac{4}{3}\sum_{i=2}^{n} x_1^3 - \left(\frac{x_1 - \sqrt{3}x_i}{2} \right)^3 + \left(\frac{-x_1 + \sqrt{3}x_i}{2} \right)^3,$$

where we applied (3.10) to each binary Chebyshev form $\text{Ч}_{3,2}(x_1, x_i) = x_1^3 - 3x_1 x_i^2$. The obtained formula is equivalent to the asserted one (1.22). □

3.6. Local minimality of cubic Chebyshev forms. This subsection is devoted to the proof of Theorem 1.3, which states that the cubic Chebyshev form $\text{Ч}_{3,n}(x) = x_1^3 - 3x_1(x_2^2 + \cdots + x_n^2)$ is a local minimum for the ratio of uniform and Bombieri norms. We denote by $G \simeq O(n-1) \subset O(n)$ the subgroup consisting of orthogonal transformations that preserve the point $(1, 0, \ldots, 0) \in \mathbb{R}^n$. Note that $G \subset O(n)$ is of codimension $n-1$ and that $\text{Ч}_{3,n}$ is invariant under G. In particular, the $O(n)$-orbit of $\text{Ч}_{3,n}$ is at most $(n-1)$-dimensional. In the following lemma we describe the tangent space to this orbit, a result that we need for the proof of Theorem 1.3.

LEMMA 3.3. *The $O(n)$-orbit of $\text{Ч}_{3,n}$ has dimension $n-1$ and its tangent space at $\text{Ч}_{3,n}$ consists of all reducible cubics of the form $\ell \cdot q$, where ℓ is a linear form that vanishes at $(1, 0, \ldots, 0) \in \mathbb{R}^n$ and $q(x) = 3x_1^2 - x_2^2 - \cdots - x_n^2$.*

Proof. For $i = 2, \ldots, n$ let us consider the elementary rotation $R_i(\varphi) \in O(n)$ in the (x_1, x_i)-plane, that is, $R_i(\varphi)$ is given by the $n \times n$ matrix whose only nonzero entries are $(R_i(\varphi))_{11} = (R_i(\varphi))_{ii} = \cos(\varphi)$, $(R_i(\varphi))_{1i} = -(R_i(\varphi))_{i1} = \sin\varphi$, and $(R_i(\varphi))_{jj} = 1$ for $j \neq 1, i$. It is a straightforward calculation to check that the tangent vector to the curve $\varphi \mapsto R_i(\varphi)^* \text{Ч}_{3,n}$ at $\varphi = 0$ is a nonzero cubic proportional to $x_i q$. For $i = 2, \ldots, n$ these $n-1$ tangent vectors are linearly independent, and, since the $O(n)$-orbit of $\text{Ч}_{3,n}$ is at most $(n-1)$-dimensional, the claim follows. □

Proof of Theorem 1.3. Let $S = \{p \in P_{3,n} : \|p\|_B = \|\text{Ч}_{3,n}\|_B\}$ denote the sphere of radius $\|\text{Ч}_{3,n}\|_B$ in $(P_{3,n}, \|\cdot\|_B)$. Denote by H any $(n-1)$-dimensional submanifold of $O(n)$ that passes through the identity $\text{id} \in O(n)$ and intersects G transversally at $\text{id} \in H \cap G$. Denote also by M any submanifold of S that has codimension $n-1$, passes through $\text{Ч}_{3,n} \in S$, and intersects the $O(n)$-orbit of $\text{Ч}_{3,n}$ transversally at $\text{Ч}_{3,n}$. Consider now the smooth map $f : H \times M \to S$, $(h, m) \mapsto h^* m$ and note that by construction the differential of f at $(\text{id}, \text{Ч}_{3,n})$ is surjective. In particular, f maps some open neighborhood of $(\text{id}, \text{Ч}_{3,n}) \in H \times M$ to an open neighborhood of $\text{Ч}_{3,n} \in S$. Therefore, since the uniform norm is $O(n)$-invariant, in order to prove the claim of the theorem it is enough to show that $\text{Ч}_{3,n} \in M$ is a local minimum of the uniform norm restricted to M. To prove the latter let us denote by T the sphere of radius $\|\text{Ч}_{3,n}\|_B$ in the tangent space to M at $\text{Ч}_{3,n}$. We claim that there exists a constant $\delta > 0$ such that for any $p \in T$ there exists a point $x \in \mathbb{R}^n$, $\|x\| = 1$, such that

(3.11) $|\text{Ч}_{3,n}(x)| = 1$ and $\text{Ч}_{3,n}(x)p(x) \geq \delta.$

It then follows for the geodesic $\gamma_p(t) = \cos t \cdot \text{Ч}_{3,n} + \sin t \cdot p$ that

$$\|\gamma_p(t)\|_\infty \geq |\cos t\, \text{Ч}_{3,n}(x) + \sin t\, p(x)| \geq \cos t + \delta \sin t \geq 1 = \|\text{Ч}_{3,n}\|_\infty$$

for all $0 \leq t \leq t_\delta$, where $t_\delta > 0$ depends only on δ. This proves that $\text{Ч}_{3,n} \in M$ is a local minimum of the uniform norm restricted to M.

In order to show (3.11) let us define

$$C_n = \{\pm e_1\} \cup \left\{ \pm\frac{1}{2}e_1 + \frac{\sqrt{3}}{2}\rho e_2 : \rho \in G \right\}$$
$$= \{\pm e_1\} \cup \{x \in \mathbb{R}^n : \|x\| = 1, \, 3x_1^2 - x_2^2 - \cdots - x_n^2 = 0\},$$

where $e_1 = (1, 0, \ldots, 0)$ and $e_2 = (0, 1, 0, \ldots, 0)$. From the G-invariance of $Ч_{3,n}$ one can see that C_n is the set of unit vectors $x \in \mathbb{R}^n$, $\|x\| = 1$, satisfying $|Ч_{3,n}(x)| = 1$ and

$$(3.12) \qquad Ч_{3,n}(\pm e_1) = \pm 1, \quad Ч_{3,n}\left(\pm\frac{1}{2}e_1 + \frac{\sqrt{3}}{2}\rho e_2\right) = \mp 1, \quad \rho \in G.$$

From Lemma 3.3 it follows that a nonzero form $p \in P_{3,n}$ vanishes on C_n if and only if it belongs to the tangent space of the $O(n)$-orbit of $Ч_{3,n}$ at $Ч_{3,n}$.[2] In particular, no $p \in T$ vanishes on the whole of C_n. From compactness of both C_n and T we hence conclude that

$$\max_{x \in C_n} |p(x)| \geq \delta'$$

for some $\delta' > 0$ and all $p \in T$. Now put $\delta = \delta'/(10n)$. Given $p \in T$, let $x' \in C_n$ be such that $|p(x')| \geq \delta'$. If $p(x')$ and $Ч_{3,n}(x')$ have the same sign, (3.11) obviously holds as $\delta' > \delta$. We now treat the case when $Ч_{3,n}(x')p(x') < 0$. Note first that, as a consequence of the G-invariance of $Ч_{3,n}$, together with the decomposition (1.22), we have the whole family of decompositions

$$(3.13) \qquad Ч_{3,n}(x) = \frac{n+2}{3}x_1^3 + \frac{4}{3}\sum_{i=2}^{n} -\langle v_{\rho,i}^+, x\rangle^3 + \langle v_{\rho,i}^-, x\rangle^3, \quad \rho \in G,$$

where $v_{\rho,i}^{\pm} = \pm 1/2 e_1 + \sqrt{3}/2\rho e_i$ and e_i is the ith unit vector, $i = 1, \ldots, n$. The set of possible $v_{\rho,i}^{\pm}$ for different $\rho \in G$ coincides with $C_n \setminus \{\pm e_1\}$. Therefore, x' either is $\pm e_1$ (in which case we can assume that $x' = e_1$), or is among $v_{\rho,i}^{\pm}$, $i = 2, \ldots, n$, for some $\rho \in G$. Since p is tangent to S at $Ч_{3,n}$, that is, $\langle p, Ч_{3,n}\rangle_B = 0$, and since $\langle p, \langle v, \cdot\rangle^3\rangle_B = p(v)$, we get from (3.12) and (3.13) that

$$0 = \langle p, Ч_{3,n}\rangle_B = \frac{n+2}{3}Ч_{3,n}(e_1)p(e_1) + \frac{4}{3}\sum_{i=2}^{n} Ч_{3,n}(v_{\rho,i}^+)p(v_{\rho,i}^+) + Ч_{3,n}(v_{\rho,i}^-)p(v_{\rho,i}^-).$$

One of these terms features $Ч_{3,n}(x')p(x') \leq -\delta'$. Elementary estimates then show that for some x among e_1 and $v_{\rho,i}^+, v_{\rho,i}^-$, $i = 2, \ldots, n$, we must have $Ч_{3,n}(x)p(x) \geq \delta'/(10n) = \delta$. We thus have verified (3.11) for all $p \in T$ and some $\delta > 0$, which concludes the proof. □

---

[2]It is interesting to state this in the language of symmetric tensors: the tangent space of the $O(n)$-orbit of the symmetric tensor associated with $Ч_{3,n}$ is the orthogonal complement of the span of its symmetric best rank-one approximations.

REFERENCES

[1] S. Banach, *Über homogene Polynome in* (L^2), Stud. Math., 7 (1938), pp. 36–44.

[2] B. Beauzamy, E. Bombieri, P. Enflo, and H. L. Montgomery, *Products of polynomials in many variables*, J. Number Theory, 36 (1990), pp. 219–245.

[3] P. L. Chebyshev, *Théorie des mécanismes connus sous le nom de parallélogrammes*, Mém. Acad. Sci. Pétersb., 7 (1854), pp. 539–568.

[4] F. H. Clarke, *Generalized gradients and applications*, Trans. Amer. Math. Soc., 205 (1975), pp. 247–262.

[5] F. H. Clarke, *Optimization and Nonsmooth Analysis*, Classics in Appl. Math. 5, SIAM, Philadelphia, PA, 1990.

[6] F. Cobos, T. Kühn, and J. Peetre, *Schatten-von Neumann classes of multilinear forms*, Duke Math. J., 65 (1992), pp. 121–156.

[7] F. Cobos, T. Kühn, and J. Peetre, *On* $\mathfrak{G}_p$-*classes of trilinear forms*, J. London Math. Soc. (2), 59 (1999), pp. 1003–1022.

[8] H. Derksen, S. Friedland, L.-H. Lim, and L. Wang, *Theoretical and Computational Aspects of Entanglement*, arXiv:1705.07160, 2017.

[9] A. Franc, *Etude Algèbrique des Multitableaux: Apports de l'Algèbre Tensorielle*, Ph.D. thesis, Université de Montpellier II, Montpellier, France, 1992.

[10] S. Friedland and L.-H. Lim, *Nuclear norm of higher-order tensors*, Math. Comp., 87 (2018), pp. 1255–1281.

[11] V. L. Goncharov, *The theory of best approximation of functions*, J. Approx. Theory, 106 (2000), pp. 2–57.

[12] G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities*, 2nd ed., Cambridge University Press, Cambridge, UK, 1952.

[13] X. Kong and D. Meng, *The bounds for the best rank-1 approximation ratio of a finite dimensional tensor space*, Pac. J. Optim., 11 (2015), pp. 323–337.

[14] T. Kühn and J. Peetre, *Embedding constants of trilinear Schatten-von Neumann classes*, Proc. Est. Acad. Sci. Phys. Math., 55 (2006), pp. 174–181.

[15] Z. Li, Y. Nakatsukasa, T. Soma, and A. Uschmajew, *On orthogonal tensors and best rank-one approximation ratio*, SIAM J. Matrix Anal. Appl., 39 (2018), pp. 400–425.

[16] Z. Li and Y.-B. Zhao, *On norm compression inequalities for partitioned block tensors*, Calcolo, 57 (2020), 11.

[17] N. N. Osipov and N. S. Sazhin, *An extremal property of Chebyshev polynomials*, Russian J. Numer. Anal. Math. Model., 23 (2008), pp. 89–95.

[18] L. Qi, *The best rank-one approximation ratio of a tensor space*, SIAM J. Matrix Anal. Appl., 32 (2011), pp. 430–442.

[19] V. Temlyakov, *Greedy Approximation*, Cambridge University Press, Cambridge, UK, 2011.

[20] A. Uschmajew, *Some results concerning rank-one truncated steepest descent directions in tensor spaces*, in Proceedings of the International Conference on Sampling Theory and Applications, 2015, pp. 415–419.