TENSOR THETA NORMS AND LOW RANK RECOVERY

HOLGER RAUHUT AND ŽELJKA STOJANAC

Abstract. We study extensions of compressive sensing and low rank matrix recovery to the recovery of tensors of low rank from incomplete linear information. While the reconstruction of low rank matrices via nuclear norm minimization is rather well-understand by now, almost no theory is available so far for the extension to higher order tensors due to various theoretical and computational difficulties arising for tensor decompositions. In fact, nuclear norm minimization for matrix recovery is a tractable convex relaxation approach, but the extension of the nuclear norm to tensors is NP-hard to compute. In this article, we introduce convex relaxations of the tensor nuclear norm which are computable in polynomial time via semidefinite programming. Our approach is based on theta bodies, a concept from computational algebraic geometry which is similar to the one of the better known Lasserre relaxations. We introduce polynomial ideals which are generated by the second order minors corresponding to different matricizations of the tensor (where the tensor entries are treated as variables) such that the nuclear norm ball is the convex hull of the algebraic variety of the ideal. The theta body of order k for such an ideal generates a new norm which we call the θ_k-norm. We show that in the matrix case, these norms reduce to the standard nuclear norm. For tensors of order three or higher however, we indeed obtain new norms. By providing the Gröbner basis for the ideals, we explicitly give semidefinite programs for the computation of the θ_k-norm and for the minimization of the θ_k-norm under an affine constraint. Finally, numerical experiments for order three tensor recovery via θ_1-norm minimization suggest that our approach successfully reconstructs tensors of low rank from incomplete linear (random) measurements.

Keywords: low rank tensor recovery, tensor nuclear norm, theta bodies, compressive sensing, semidefinite programming, convex relaxation, polynomial ideals, Gröbner bases

MSC 2010: 13P10, 15A69, 15A60, 52A41, 90C22, 90C25, 94A20,

1. Introduction and Motivation

Compressive sensing predicts that sparse vectors can be recovered from underdetermined linear measurements via efficient methods such as ℓ_1-minimization [9, 15, 18]. This finding has various applications in signal and image processing and beyond. It has recently been observed that the principles of this theory can be transferred to the problem of recovering a low rank matrix from underdetermined linear measurements. One prominent choice of recovery method consists in minimizing the nuclear norm subject to the given linear constraint [17, 47]. This convex optimization problem can be solved efficiently and recovery results for certain random measurement maps have been provided, which quantify the minimal number of measurements required for successful recovery [47, 6, 5, 27, 26, 36].

There is significant interest in going one step further and to extend the theory to the recovery of low rank tensors (higher-dimensional arrays) from incomplete linear measurements. Applications include image and video inpainting [38], reflectance data recovery [38] (e.g. for use in photo-realistic raytracers), machine learning [48] and seismic data processing [34]. Several approaches have already been introduced [38, 20, 31, 44, 45], but unfortunately, so far, for none of them a completely satisfactory theory is available. Either the method is not tractable [51], or no (complete) rigorous recovery results quantifying the minimal number of measurements are available [20, 38, 44, 45, 32, 14, 35], or the available bounds are highly nonoptimal [31, 16, 39]. For instance, a version of the nuclear norm for higher order tensors can be introduced, but unfortunately, its computation (and therefore, also its minimization) is NP-hard [19] – nevertheless, some recovery results for tensor completion via nuclear norm minimization are available in [51]. Moreover,

Date: May 19, 2015.

versions of iterative hard thresholding for various tensor formats have been introduced [44, 45]. This approach leads to a computationally tractable algorithm, which empirically works well. However, only a partial analysis based on the restricted isometry property has been provided, which so far only shows convergence under a condition on the iterates that cannot be checked a priori. Nevertheless, the restricted isometry property (RIP) has been analyzed for certain random measurement maps [44, 45, 46]. These near optimal bounds on the number of measurements ensuring the RIP, however, provide only a hint on how many measurements are required because the link between the RIP and recovery is so far only partial [45, 46].

This article introduces a new approach for tensor recovery based on convex relaxation. The idea is to further relax the nuclear norm in order to arrive at a norm which can be computed (and minimized under a linear constraint) in polynomial time. The hope is that the new norm is only a slight relaxation and possesses very similar properties as the nuclear norm. Our approach is based on so-called theta bodies, a concept from computational algebraic geometry [40, 22, 2] which is similar to the better known Lasserre relaxations [37]. We arrive at a whole family of convex bodies (indexed by a polynomial degree), which form convex relaxations of the nuclear norm ball. The resulting norms are called theta norms. The corresponding unit norm balls are nested and contain the nuclear norm ball. They can be computed by semidefinite optimization, and also the minimization of the θ_k norm subject to a linear constraints is a semidefinite program (SDP) whose solution can be computed in polynomial time – the complexity growing with k.

The basic idea for the construction of these new norms is to define polynomial ideals, where each variable corresponds to an entry of the tensor, such that its algebraic variety consists of the rank one tensors of Frobenius norm one. The convex hull of this set is the tensor nuclear norm ball. The ideals that we propose are generated by the minors of order two of all matricizations of the tensor (or at least of a subset of the possible matricizations) together with the polynomial corresponding to the squared Frobenius norm minus one. Here, a matricization denotes a matrix which is generated from the tensor by combining several indices to a row index, and the remaining indices to a column index. In fact, all such minors being zero simultaneously means that the tensor has rank one. The k-theta body of the ideal corresponds then to a relaxation of the convex hull of its algebraic variety, i.e., to a further relaxation of the tensor nuclear norm. The index $k \in \mathbb{N}$ corresponds to a polynomial degree involved in the construction of the theta bodies (some polynomial is required to be k-sos modulo the ideal, see below), and $k = 1$ leads to the largest theta body in a family of convex relaxations.

We will show that for the matrix case (tensors of order 2), our approach does not lead to new norms. All resulting theta norms are rather equal to the matrix nuclear norm. This fact suggests that the theta norms in the higher order tensor case are all natural generalizations of the matrix nuclear norm.

We derive the corresponding semidefinite programs explicitly and present numerical experiments which show that θ_1-norm minimization successfully recovers tensors of low row rank from few random linear measurements. Unfortunately, a rigorous theoretical analysis of the recovery performance of TH_k-minimization is not yet available but will be the subject of future studies.

1.1. **Low rank matrix recovery.** Before passing to tensor recovery, we recall some basics on matrix recovery. Let $\mathbf{X} \in \mathbf{R}^{n_1 \times n_2}$ of rank at most $r \ll \min\{n_1, n_2\}$, and suppose we are given linear measurements

$$\mathbf{y} = \mathcal{A}(\mathbf{X}),$$

where $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^m$ is a linear map with $m \ll n_1 n_2$. Reconstructing $\mathbf{X}$ from $\mathbf{y}$ amounts to solving an underdetermined linear system. Unfortunately, the rank minimization problem of computing the minimizer of

$$\min_{\mathbf{Z} \in \mathbb{R}^{n_1 \times n_2}} \operatorname{rank}(\mathbf{Z}) \quad \text{subject to } \mathcal{A}(\mathbf{Z}) = \mathbf{y}$$

is NP-hard in general. As a tractable alternative, the convex optimization problem

$$\min_{\mathbf{Z} \in \mathbb{R}^{n_1 \times n_2}} \|\mathbf{Z}\|_* \quad \text{subject to } \mathcal{A}(\mathbf{Z}) = \mathbf{y} \tag{1}$$

has been suggested [17, 47], where the nuclear norm $\|\cdot\|_* = \sum_j \sigma_j(\mathbf{Z})$ is the sum of the singular values of $\mathbf{Z}$. This problem can be solved efficiently by various methods [3]. For instance, it can be reformulated as a semidefinite program [17], but splitting methods may be more efficient [43, 49].

While it is hard to analyze low rank matrix recovery for deterministic measurement maps, optimal bounds are available for several random matrix constructions. If $\mathcal{A}$ is a Gaussian measurement map, i.e.,

$$\mathcal{A}(\mathbf{X})_j = \sum_{k,\ell} \mathcal{A}_{jk\ell} X_{k\ell}, \quad j \in [m] := \{1, 2, \ldots, m\},$$

where the $\mathcal{A}_{jkl}$, $j \in [m], k \in [n_1], \ell \in [n_2]$, are independent mean-zero, variance one Gaussian random variables, then a matrix $\mathbf{X}$ of rank at most r can be reconstructed exactly from $\mathbf{y} = \mathcal{A}(\mathbf{X})$ via nuclear norm minimization (1) with probability at least $1 - e^{-cm}$ provided that

$$m \geq Crn, \qquad n = \max\{n_1, n_2\}, \tag{2}$$

where the constants $c, C > 0$ are universal [5, 10]. Moreover, the reconstruction is stable under passing to only approximately low rank matrices and under adding noise on the measurements. Another interesting measurement map corresponds to the matrix completion problem [6, 8, 26, 11], where the measurements are randomly chosen entries of the matrix $\mathbf{X}$. Measurements taken as Frobenius inner products with rank-one matrices are studied in [36], and arise in the phase retrieval problem as special case [7]. Also here, $m \geq Crn$ (or $m \geq Crn\log(n)$ for certain structured measurements) is sufficient for exact recovery.

1.2. **Tensor recovery.** An order d-tensor (or mode-d-tensor) is an element $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ indexed by $[n_1] \times [n_2] \times \cdots \times [n_d]$. Of course, the case $d = 2$ corresponds to matrices. For $d \geq 3$, several notions and computational tasks become much more involved than for the matrix case. Already the notion of rank requires some clarification, and in fact, several different definitions are available, see for instance [33, 25]. We will mainly work with the so-called canonical rank or CP-rank in the following. A dth order tensor $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ is of rank one if there exist vectors $\mathbf{u}^1 \in \mathbb{R}^{n_1}, \mathbf{u}^2 \in \mathbb{R}^{n_2}, \ldots, \mathbf{u}^d \in \mathbb{R}^{n_d}$ such that $\mathbf{X} = \mathbf{u}^1 \otimes \mathbf{u}^2 \otimes \cdots \otimes \mathbf{u}^d$ or elementwise

$$X_{i_1 i_2 \ldots i_d} = u^1_{i_1} u^2_{i_2} \cdots u^d_{i_d}.$$

The CP-rank (or canonical rank and in the following just rank) of a tensor $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$, similarly as in the matrix case, is the smallest number of rank one tensors that sum up to $\mathbf{X}$.

Given a linear measurement map $\mathcal{A} : \mathbb{R}^{n_1 \times \cdots \times n_d} \to \mathbb{R}^m$ (which can represented as a $(d+1)$th order tensor), our aim is to recover a tensor $\mathbf{X} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ from $\mathbf{y} = \mathcal{A}(\mathbf{X})$ when $m \ll n_1 \cdot n_2 \cdots n_d$. The matrix case $d = 2$ suggests to consider minimization of the tensor nuclear norm for this task,

$$\min_{\mathbf{Z}} \|\mathbf{Z}\|_* \quad \text{subject to } \mathcal{A}(\mathbf{Z}) = \mathbf{y},$$

where the nuclear norm is defined as

$$\|\mathbf{X}\|_* = \min \left\{ \sum_{k=1}^{r} |c_k| : \mathbf{X} = \sum_{k=1}^{r} c_k \mathbf{u}^{1,k} \otimes \mathbf{u}^{2,k} \otimes \cdots \otimes \mathbf{u}^{d,k}, r \in \mathbb{N}, \left\|\mathbf{u}^{i,k}\right\|_{\ell_2} = 1, i \in [d], k \in [r] \right\}.$$

Unfortunately, in the tensor case, computing the canonical rank of a tensor, as well as computing the nuclear norm of a tensor is NP-hard in general, see [30, 29, 19]. Let us nevertheless mention that some theoretical results for tensor recovery via nuclear norm minimization are contained in [51].

The aim of this article is to introduce relaxations of the tensor nuclear norm, based on so-called theta bodies, which is both computationally tractable and whose minimization allows for exact recovery of low rank tensors from incomplete linear measurements.

Let us remark that one may reorganize (flatten) a low rank tensor $\mathbf{X} \in \mathbb{R}^{n \times n \times n}$ into a low rank matrix $\tilde{\mathbf{X}} \in \mathbb{R}^{n \times n^2}$ and simply apply concepts from matrix recovery. However, the bound (2) on the required number of measurements then reads

$$m \geq Crn^2. \tag{3}$$

Moreover, it has been suggested in [20, 50, 38] to minimize the sum of nuclear norms of the unfoldings (different reorganizations of the tensor as a matrix) subject to the linear constraint matching the measurements. Although this seems to be a reasonable approach at first sight, it has been shown in [42], that this approach cannot work with less measurements than stated by the estimate in (3).

Bounds for a version of the restricted isometry property for certain tensor formats in [46] suggest that

$$m \geq Cr^2 n$$

measurements should be sufficient when working directly with the tensor structure – precisely, this bound uses the so-called tensor train format [41]. (Possibly, the term r^2 may even be lowered to r when using the "right" tensor format.) However, connecting the restricted isometry property in a completely satisfactory way with the success of an efficient tensor recovery algorithm is still open. (Partial results are contained in [46].) In any case, this suggests that one should exploit the tensor structure of the problem rather than reducing to a matrix recovery problem in order to recover a low rank tensor using the minimal number of measurements. Of course, similar considerations apply to tensors of order higher than three, where the difference between the reduction to the matrix case and working directly with the tensor structure will become even stronger.

1.3. **Some notation.** We write vectors with small bold letters, matrices and tensors with capital bold letters and sets with capital calligraphic letters. The cardinality of a set $\mathcal{S}$ is denoted by $|\mathcal{S}|$.

For a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ and subsets $\mathcal{I} \subset [m]$, $\mathcal{J} \subset [n]$ the submatrix of $\mathbf{A}$ with columns indexed by $\mathcal{I}$ and rows indexed by $\mathcal{J}$ is denoted by $\mathbf{A}_{\mathcal{I},\mathcal{J}}$. A set of all minors of $\mathbf{A}$ order k is of the form

$$\{\det(\mathbf{A}_{\mathcal{I},\mathcal{J}}) : \mathcal{I} \subset [m]\,, \mathcal{J} \subset [n]\,, |\mathcal{I}| = |\mathcal{J}| = k\}\,.$$

The Frobenius norm of a matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$ is given as

$$\|\mathbf{X}\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} X_{ij}^2} = \sqrt{\sum_{i=1}^{\min\{m,n\}} \sigma_i^2},$$

where the σ_i list the singular values of $\mathbf{X}$. The nuclear norm is given by $\|\mathbf{X}\|_* = \sum_{i=1}^{\min\{m,n\}} \sigma_i$. It is well-known that its unit ball is the convex hull of all rank one matrices of Frobenius norm one.

The vectorization of a tensor $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ is denoted by $\mathrm{vec}(\mathbf{X}) \in \mathbb{R}^{n_1 n_2 \cdots n_d}$. The ordering of the elements in $\mathrm{vec}(\mathbf{X})$ is not important as long as it remains consistent. Fibers are a higher order analogue of matrix rows and columns. For $k \in [d]$, the mode-k fiber of a dth order tensor is obtained by fixing every index except for the k-th one. The Frobenius norm of a d-th order tensor $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ is defined as

$$\|\mathbf{X}\|_F = \sqrt{\sum_{i_1=1}^{n_1}\sum_{i_2=1}^{n_2} \cdots \sum_{i_d=1}^{n_d} X_{i_1 i_2 \cdots i_d}^2}.$$

Matricization (also called flattening) is the operation that transforms a tensor into a matrix. More precisely, for a dth order tensor $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ and an ordered subset $\mathcal{S} \subseteq [d]$, an $\mathcal{S}$-matricization $\mathbf{X}^{\mathcal{S}} \in \mathbb{R}^{\prod_{i \in \mathcal{S}} n_i \times \prod_{k \in \mathcal{S}^c} n_k}$ is defined as

$$X^{\mathcal{S}}_{(i_k)_{k \in \mathcal{S}}, (i_\ell)_{\ell \in \mathcal{S}^c}} = X_{i_1, i_2, \ldots, i_d},$$

i.e., the indexes in the set $\mathcal{S}$ define the rows of a matrix and the indexes in the set $\mathcal{S}^c = [d] \setminus \mathcal{S}$ define the columns. For a singelton set $\mathcal{S} = \{i\}$, for $i \in [d]$, we call the $\mathcal{S}$-matricization the i-th unfolding. Notice that every $\mathcal{S}$-matricization of a rank one tensor is a rank one matrix. Conversely, if every $\mathcal{S}$-matricization of a tensor is a rank one matrix, then the tensor is of rank one. This is even true, if all unfoldings of a tensor are of rank one.

We often use MATLAB notation. Specifically, for a dth order tensor $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$, we write $\mathbf{X}(:, :, \ldots, :, k)$ for the $(d-1)$-order subtensor in $\mathbb{R}^{n_1 \times n_2 \times \cdots \times n_{d-1}}$ obtained by fixing the last index i_d to k. For simplicity, the subscripts $i_1 i_2 \cdots i_d$ and $\hat{i}_1 \hat{i}_2 \cdots \hat{i}_d$ will often be denoted by I and $\hat{I}$, respectively. In particular, instead of writing $X_{i_1 i_2 \dots i_d} X_{\hat{i}_1 \hat{i}_2 \dots \hat{i}_d}$, we often just write $X_I X_{\hat{I}}$. Below, we will use the grevlex ordering of monomials indexed by subscripts I, which in particular requires to define an ordering for such subscripts. We make the agreement that $I > \hat{I}$ if $i_1 > \hat{i}_1$ or if $i_1 = \hat{i}_1, i_2 = \hat{i}_2, \dots i_{\ell-1} = \hat{i}_{\ell-1}$ and $i_\ell > \hat{i}_\ell$ for some $\ell \in [d]$.

1.4. **Structure of the paper.** In Section 2 we will review the basic definition and properties of theta bodies. Section 3 considers the matrix case. We introduce a suitable polynomial ideal whose algebraic variety is the set of rank one Frobenius norm one matrices. We discuss the corresponding θ_k-norms and show that they all coincide with the matrix nuclear norm. The case of 2×2-matrices is described in detail. In Section 4 we pass to the tensor case and discuss first the case of order three tensors. We introduce a suitable polynomial ideal, provide its reduced Gröbner basis and define the corresponding θ_k-norms. Then we pass to order four tensors, where it turns out that one may define different ideals (corresponding to different sets of matricizations) whose algebraic variety is the set of rank one Frobenius norm one tensors. These different ideals lead to different θ_k-norms, and we discuss two possibilities in detail. The general d-th order case is discussed at the end of Section 4. Here, we concentrate only on one possibility for the polynomial ideal which corresponds to the set of all possible matricizations of the tensor. We show that a certain set of order two minors forms the Gröbner basis for this ideal, which is key for defining the θ_k-norms. Section 5 briefly discusses the polynomial runtime of the algorithms for computing and minimizing the θ_k-norms showing that our approach is tractable. Numerical experiments for low rank recovery of third order tensors are presented in Section 6, which show that our approach successfully recovers low rank tensor from incomplete Gaussian random measurements. Appendix 7 discusses some background from algebraic geometry (monomial orderings and Gröbner bases) that is required throughout the main body of the article.

2. Theta bodies

As outlined above, we will introduce new tensor norms as relaxations of the nuclear norm in order to come up with a new convex optimization approach for low rank tensor recovery. Our approach builds on the concept of theta bodies, a recent concept from computational algebraic geometry, which is similar to Lasserre relaxations [37]. In order to introduce it, we first discuss the necessary basics from algebraic geometry. For more information, we refer to [12, 13] and to the appendix.

For a non-zero polynomial $f = \sum_{\boldsymbol{\alpha}} a_{\boldsymbol{\alpha}} \mathbf{x}^{\boldsymbol{\alpha}}$ in $\mathbb{R}[\mathbf{x}] = \mathbb{R}[x_1, x_2, \dots, x_n]$ and a monomial order $>$, we denote

a) the multidegree of f by $\text{multideg}(f) = \max\left(\boldsymbol{\alpha} \in \mathbb{Z}^n_{\geq 0} : a_{\boldsymbol{\alpha}} \neq 0\right)$,
b) the leading coefficient of f by $\text{LC}(f) = a_{\text{multideg}(f)} \in \mathbb{R}$,
c) the leading monomial of f by $\text{LM}(f) = \mathbf{x}^{\text{multideg}(f)}$,
d) the leading term of f by $\text{LT}(f) = \text{LC}(f)\,\text{LM}(f)$.

Let $J \subset \mathbb{R}[\mathbf{x}]$ be a polynomial ideal. Its real algebraic variety is the set of all points in $\mathbf{x} \in \mathbb{R}^n$ where all polynomials in the ideal vanish, i.e.,

$$\nu_{\mathbb{R}}(J) = \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) = 0, \text{ for all } f \in J\}.$$

By Hilbert's basis theorem [13] every polynomial ideal in $\mathbb{R}[\mathbf{x}]$ has a finite generating set. Thus, we may assume that J is generated by a set $\mathcal{F} = \{f_1, f_2, \dots, f_k\}$ of polynomials in $\mathbb{R}[\mathbf{x}]$ and write

$$J = \langle f_1, f_2, \dots, f_k \rangle = \left\langle \{f_i\}_{i \in [k]} \right\rangle \quad \text{or simply} \quad J = \langle \mathcal{F} \rangle.$$

Its real algebraic variety is the set

$$\nu_{\mathbb{R}}(J) = \{\mathbf{x} \in \mathbb{R}^n : f_i(\mathbf{x}) = 0 \text{ for all } i \in [k]\}.$$

Throughout the paper, $\mathbb{R}[\mathbf{x}]_k$ denotes the set of polynomials of degree at most k. A degree one polynomial is also called linear polynomial. A very useful certificate for positivity of polynomials is contained in the following definition [22].

Definition 1. Let J be an ideal in $\mathbb{R}[\mathbf{x}]$. A polynomial $f \in \mathbb{R}[\mathbf{x}]$ is *k-sos mod J* if there exists a finite set of polynomials $h_1, h_2, \ldots, h_t \in \mathbb{R}[\mathbf{x}]_k$ such that $f \equiv \sum_{j=1}^t h_j^2 \bmod J$, i.e, if $f - \sum_{j=1}^t h_j^2 \in J$.

A special case of theta bodies was first introduced by Lovász in [40] and in full generality they appeared in [22]. Later, they have been analyzed in [21, 23]. The definitions and theorems in the remainder of the section are taken from [22].

Definition 2 (Theta body)**.** Let $J \subseteq \mathbb{R}[\mathbf{x}]$ be an ideal. For a positive integer k, the *k-th theta body of J* is defined as

$$\mathrm{TH}_k(J) := \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) \geq 0 \text{ for every linear } f \text{ that is } k\text{-sos mod } J\}.$$

We say that an ideal $J \subseteq \mathbb{R}[\mathbf{x}]$ is TH_k*-exact* if $\mathrm{TH}_k(J)$ equals $\overline{\mathrm{conv}(\nu_{\mathbb{R}}(J))}$, the closure of the convex hull of $\nu_{\mathbb{R}}(J)$.

Theta bodies are closed convex sets, while $\mathrm{conv}(\nu_{\mathbb{R}}(J))$ may not necessarily be closed and by definition,

$$\mathrm{TH}_1(J) \supseteq \mathrm{TH}_2(J) \supseteq \cdots \supseteq \mathrm{conv}(\nu_{\mathbb{R}}(J)). \tag{4}$$

The theta-body sequence of J can converge (finitely or asymptotically), if at all, only to $\overline{\mathrm{conv}(\nu_{\mathbb{R}}(J))}$. More on guarantees on convergence can be found in [22, 23]. However, to our knowledge, none of the existing guarantees apply to the cases discussed below.

Given any polynomial, it is possible to check whether it is k-sos mod J using a Gröbner basis and semidefinite programming. However, using this definition in practice requires knowledge of all linear polynomials (possibly infinitely many) that are k-sos mod J. To overcome this difficulty, we need an alternative description of $\mathrm{TH}_k(J)$ discussed next.

As in [2], we assume that there are no linear polynomials in the ideal J. Otherwise, some variable x_i would be congruent to a linear combination of other variables modulo J and we could work in a smaller polynomial ring $\mathbb{R}[\mathbf{x}^i] = \mathbb{R}[x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n]$. Therefore, $\mathbb{R}[\mathbf{x}]_1 / J \cong \mathbb{R}[\mathbf{x}]_1$ and $\{1 + J, x_1 + J, \ldots, x_n + J\}$ can be completed to a basis $\mathcal{B}$ of $\mathbb{R}[\mathbf{x}]/J$. Recall that the degree of an equivalence class $f + J$, denoted by $\deg(f + J)$, is the smallest degree of an element in the class. We assume that each element in the basis $\mathcal{B} = \{f_i + J\}$ of $\mathbb{R}[\mathbf{x}]/J$ is represented by the polynomial whose degree equals the degree of its equivalence class, i.e., $\deg(f_i + J) = \deg(f_i)$. In addition, we assume that $\mathcal{B} = \{f_i + J\}$ is ordered so that $f_{i+1} > f_i$, where $>$ is a fixed monomial ordering. Further, we define the set $\mathcal{B}_k$

$$\mathcal{B}_k := \{f + J \in \mathcal{B} : \deg(f + J) \leq k\}.$$

Definition 3 (Theta basis)**.** Let $J \subseteq \mathbb{R}[\mathbf{x}]$ be an ideal. A basis $\mathcal{B} = \{f_0 + J, f_1 + J, \ldots\}$ of $\mathbb{R}[\mathbf{x}]/J$ is a θ-basis if it has the following properties

1) $\mathcal{B}_1 = \{1 + J, x_1 + J, \ldots, x_n + J\}$,
2) if $\deg(f_i + J), \deg(f_j + J) \leq k$ then $f_i f_j + J$ is in the $\mathbb{R}$-span of $\mathcal{B}_{2k}$.

As in [2, 22] we consider only monomial bases $\mathcal{B}$ of $\mathbb{R}[\mathbf{x}]/J$, i.e., bases $\mathcal{B}$ such that f_i is a monomial, for all $f_i + J \in \mathcal{B}$.

For determining a θ-basis, we first need to compute the reduced Gröbner basis $\mathcal{G}$ of the ideal J, see Definitions 7 and 8. The set $\mathcal{B}$ will satisfy the second property in the definition of the theta basis if the reduced Gröbner basis is with respect to an ordering which first compares the total degree. Therefore, throughout the paper we use the graded reverse monomial ordering (Definition 6) or simply grevlex ordering, although also the graded lexicographic ordering would be appropriate.

A technique to compute a θ-basis $\mathcal{B}$ of $\mathbb{R}[\mathbf{x}]/J$ consists in taking $\mathcal{B}$ to be the set of equivalence classes of the standard monomials of the corresponding initial ideal

$$J_{\text{initial}} = \left\langle \{\mathrm{LT}(f)\}_{f \in J} \right\rangle = \left\langle \{\mathrm{LT}(g_i)\}_{i \in [s]} \right\rangle,$$

where $\mathcal{G} = \langle g_1, g_2, \ldots, g_s \rangle$ is the reduced Gröbner basis of the ideal J. In other words, a set $\mathcal{B} = \{f_0 + J, f_1 + J, \ldots\}$ will be a θ-basis of $\mathbb{R}[\mathbf{x}]/J$ if it contains all $f_i + J$ such that

1) f_i is a monomial
2) f_i is not divisible by any of the monomials in the set $\{\mathrm{LT}(g_i) : i \in [s]\}$.

As the next important tool we need the so-called combinatorial moment matrix of J. To this end, we fix a θ-basis $\mathcal{B} = \{f_i + J\}$ of $\mathbb{R}[\mathbf{x}]/J$ and define $[\mathbf{x}]_{\mathcal{B}_k}$ to be the column vector formed by all elements of $\mathcal{B}_k$ in order. Then $[\mathbf{x}]_{\mathcal{B}_k}[\mathbf{x}]_{\mathcal{B}_k}^T$ is a square matrix indexed by $\mathcal{B}_k$ and its (i,j)-entry is equal to $f_i f_j + J$. By hypothesis, the entries of $[\mathbf{x}]_{\mathcal{B}_k}[\mathbf{x}]_{\mathcal{B}_k}^T$ lie in the $\mathbb{R}$-span of $\mathcal{B}_{2k}$. Let $\{\lambda_{i,j}^l\}$ be the unique set of real numbers such that $f_i f_j + J = \sum_{f_l + J \in \mathcal{B}_{2k}} \lambda_{i,j}^l (f_l + J)$.

The theta bodies TH_k can be characterized via the combinatorial moment matrix as stated in the next result from [22], which will be the basis for computing and minimization the new tensor norm introduced below via semidefinite programming.

Definition 4. Let $J, \mathcal{B}$ and $\{\lambda_{i,j}^l\}$ be as above. Let $\mathbf{y}$ be a real vector indexed by $\mathcal{B}_{2k}$ with $y_0 = 1$, where y_0 is the first entry of $\mathbf{y}$, indexed by the basis element $1 + J$. The k*-th combinatorial moment matrix* $\mathbf{M}_{\mathcal{B}_k}(\mathbf{y})$ of J is the real matrix indexed by $\mathcal{B}_k$ whose (i,j)-entry is $[\mathbf{M}_{\mathcal{B}_k}(\mathbf{y})]_{i,j} = \sum_{f_l + J \in \mathcal{B}_{2k}} \lambda_{i,j}^l y_l$.

Theorem 1. *The k-th theta body of J, $\mathrm{TH}_k(J)$, is the closure of*

$$Q_{\mathcal{B}_k}(J) = \pi_{\mathbb{R}^n}\left\{\mathbf{y} \in \mathbb{R}^{\mathcal{B}_{2k}} : \mathbf{M}_{\mathcal{B}_k}(\mathbf{y}) \succeq 0, y_0 = 1\right\},$$

where $\pi_{\mathbb{R}^n}$ denotes the projection onto the variables $y_1 = y_{x_1+J}, \ldots, y_n = y_{x_n+J}$.

Table 1 shows a step-by-step procedure for computing $\mathrm{TH}_k(J)$.

Algorithm for computing $\mathrm{TH}_k(J)$

Input: An ideal $J \in \mathbb{R}[\mathbf{x}] = \mathbb{R}[x_1, x_2, \ldots, x_n]$.
Compute the reduced Gröbner basis for the ideal J
Compute a θ-basis $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2 \cup \ldots = \{f_0 + J, f_1 + J, \ldots\}$ of $\mathbb{R}[\mathbf{x}]/J$ (see Definition 3)
Compute the combinatorial moment matrix $\mathbf{M}_{\mathcal{B}_k}(\mathbf{y})$:
(1) $[\mathbf{x}]_{\mathcal{B}_k} = \{\text{all elements of } \mathcal{B}_k \text{ in order}\}$
(2) $(\mathbf{X}_{\mathcal{B}_k})_{i,j} = \left([\mathbf{x}]_{\mathcal{B}_k}[\mathbf{x}]_{\mathcal{B}_k}^T\right)_{i,j} = f_i f_j + J = \sum_{f_l + J \in \mathcal{B}_{2k}} \lambda_{i,j}^l (f_l + J)$
(3) $[\mathbf{M}_{\mathcal{B}_k}(\mathbf{y})]_{i,j} = \sum_{f_l + J \in \mathcal{B}_{2k}} \lambda_{i,j}^l y_l$
Output: $\mathrm{TH}_k(J)$ is the closure of

$$Q_{\mathcal{B}_k}(J) = \pi_{\mathbb{R}^n}\left\{\mathbf{y} \in \mathbb{R}^{\mathcal{B}_{2k}} : \mathbf{M}_{\mathcal{B}_k}(\mathbf{y}) \succeq 0, y_0 = 1\right\}.$$

TABLE 1. Algorithm for computing $\mathrm{TH}_k(J)$

3. The matrix case

As a start, we consider the matrix nuclear unit norm ball and provide hierarchical relaxations via theta bodies. The k-th relaxation defines a matrix unit θ_k-norm ball with the property

$$\|\mathbf{X}\|_{\theta_k} \leq \|\mathbf{X}\|_{\theta_{k+1}} \qquad \text{for all } \mathbf{X} \in \mathbb{R}^{m \times n} \text{ and all } k \in \mathbb{N}.$$

However, we will show that all these θ_k-norms coincide with the matrix nuclear norm.

The first step in computing hierarchical relaxations of the unit nuclear norm ball consists in finding a polynomial ideal J such that its algebraic variety (the set of points for which the ideal vanishes) coincides with the set of all rank one, Frobenius norm one matrices

$$\nu_{\mathbb{R}}(J) = \left\{\mathbf{X} \in \mathbb{R}^{m \times n} : \|\mathbf{X}\|_F = 1, \operatorname{rank}(\mathbf{X}) = 1\right\}. \tag{5}$$

Recall that the convex hull of this set is the nuclear norm ball. The following lemma states the elementary fact that a non-zero matrix is a rank one matrix if and only if all its minors of order two are zero.

For notational purposes, we define the following polynomials in $\mathbb{R}[\mathbf{x}] = \mathbb{R}[x_{11}, x_{12}, \ldots, x_{mn}]$

$$g(\mathbf{x}) = \sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij}^2 - 1 \text{ and } f_{ijkl}(\mathbf{x}) = x_{il}x_{kj} - x_{ij}x_{kl} \quad \text{for } 1 \leq i < k \leq m,\ 1 \leq j < l \leq n. \tag{6}$$

Lemma 1. *Let $\mathbf{X} \in \mathbb{R}^{m\times n}\setminus\{\mathbf{0}\}$. Then $\mathbf{X}$ is a rank one, Frobenius norm one matrix if and only if*

$$\mathbf{X} \in \mathcal{R} := \{\mathbf{X} : g(\mathbf{X}) = 0 \text{ and } f_{ijkl}(\mathbf{X}) = 0 \text{ for all } i < k, j < l\}. \tag{7}$$

Proof. If $\mathbf{X} \in \mathbb{R}^{m\times n}$ is a rank one matrix with $\|\mathbf{X}\|_F = 1$, then by definition there exist two vectors $\mathbf{u} \in \mathbb{R}^m$ and $\mathbf{v} \in \mathbb{R}^n$ such that $X_{ij} = u_i v_j$ for all $i \in [m]$, $j \in [n]$ and $\|\mathbf{u}\|_2 = \|\mathbf{v}\|_2 = 1$. Thus

$$X_{ij}X_{kl} - X_{il}X_{kj} = u_i v_j u_k v_l - u_i v_l u_k v_j = 0 \quad \text{and} \quad \sum_{i=1}^{m}\sum_{j=1}^{n} X_{ij}^2 = \sum_{i=1}^{m} u_i^2 \sum_{j=1}^{n} v_j^2 = 1.$$

For the converse, let $\mathbf{X}_{\cdot i}$ represent the i-th column of a matrix $\mathbf{X} \in \mathcal{R}$. Then, for all $j, l \in [n]$ with $j < l$, it holds

$$X_{ml} \cdot \mathbf{X}_{\cdot j} - X_{mj} \cdot \mathbf{X}_{\cdot l} = X_{ml} \cdot \begin{bmatrix} X_{1j} \\ X_{2j} \\ \vdots \\ X_{mj} \end{bmatrix} - X_{mj} \cdot \begin{bmatrix} X_{1l} \\ X_{2l} \\ \vdots \\ X_{ml} \end{bmatrix} = \begin{bmatrix} X_{1j}X_{ml} - X_{1l}X_{mj} \\ X_{2j}X_{ml} - X_{2l}X_{mj} \\ \vdots \\ X_{mj}X_{ml} - X_{mj}X_{ml} \end{bmatrix} = \mathbf{0},$$

since $X_{ij}X_{ml} = X_{il}X_{mj}$ for all $i \in [m-1]$ by definition of $\mathcal{R}$. Thus, the columns of the matrix $\mathbf{X}$ span a space of dimension one, i.e., the matrix $\mathbf{X}$ is a rank one matrix. From $\sum_{i=1}^{m}\sum_{j=1}^{n} X_{ij}^2 - 1 = 0$ it follows that the matrix $\mathbf{X}$ is normalized, i.e., $\|\mathbf{X}\|_F = 1$. $\square$

It follows from Lemma 1 that the set of rank one, Frobenius norm one matrices coincides with the algebraic variety $\nu_{\mathbb{R}}(J_{M_{mn}})$ for the ideal $J_{M_{mn}}$ generated by the polynomials g and f_{ijkl}, i.e.,

$$J_{M_{mn}} = \langle \mathcal{G}_{M_{mn}} \rangle \quad \text{with } \mathcal{G}_{M_{mn}} = \{g(\mathbf{x})\} \cup \{f_{ijkl}(\mathbf{x}) : 1 \leq i < k \leq m,\ 1 \leq j < l \leq n\}. \tag{8}$$

Recall that the convex hull of the set $\mathcal{R}$ in (7) forms the unit nuclear norm ball and by definition of the theta bodies,

$$\overline{\operatorname{conv}(\nu_{\mathbb{R}}(J_{M_{mn}}))} \subseteq \cdots \subseteq \mathrm{TH}_{k+1}(J_{M_{mn}}) \subseteq \mathrm{TH}_k(J_{M_{mn}}) \subseteq \cdots \subseteq \mathrm{TH}_1(J_{M_{mn}}).$$

Therefore, the theta bodies form closed, convex hierarchical relaxations of the matrix nuclear norm ball. In addition, the theta body $\mathrm{TH}_k(J_{M_{mn}})$ is symmetric, $\mathrm{TH}_k(J_{M_{mn}}) = -\mathrm{TH}_k(J_{M_{mn}})$. Therefore, it defines a unit ball of a norm that we call the *θ_k-norm*.

The next result shows that the generating set of the ideal $J_{M_{mn}}$ introduced above is a Gröbner basis.

Lemma 2. *The set $\mathcal{G}_{M_{mn}}$ forms the reduced Gröbner basis of the ideal $J_{M_{mn}}$ with respect to the grevlex order.*

Proof. The set $\mathcal{G}_{M_{mn}}$ is clearly a basis for the ideal $J_{M_{mn}}$. By Proposition 1 in the appendix, we only need to check whether the S-polynomial, see Definition 10, satisfies $S(p,q) \to_{\mathcal{G}_{M_{mn}}} 0$ for all $p, q \in \mathcal{G}_{M_{mn}}$ whenever the leading monomials $\mathrm{LM}(p)$ and $\mathrm{LM}(q)$ are not relatively prime. Here, $S(p,q) \to_{\mathcal{G}_{M_{mn}}} 0$ means that $S(p,q)$ reduces to 0 modulo $\mathcal{G}_{M_{mn}}$, see Definition 9.

Notice that $\mathrm{LM}(g) = x_{11}^2$ and $\mathrm{LM}(f_{ijkl}) = x_{il}x_{kj}$ are relatively prime, for all $1 \leq i < k \leq m$ and $1 \leq j < l \leq n$. Therefore, we only need to show that $S(f_{ijkl}, f_{\hat{i}\hat{j}\hat{k}\hat{l}}) \to_{\mathcal{G}_{M_{mn}}} 0$ whenever the leading monomials $\mathrm{LM}(f_{ijkl})$ and $\mathrm{LM}(f_{\hat{i}\hat{j}\hat{k}\hat{l}})$ are not relatively prime. First we consider

$$f_{ijkl}(\mathbf{x}) = x_{il}x_{kj} - x_{ij}x_{kl} \quad \text{and} \quad f_{i\hat{j}\hat{k}l}(\mathbf{x}) = x_{il}x_{\hat{k}\hat{j}} - x_{i\hat{j}}x_{\hat{k}l}$$

for $1 \leq i < k < \hat{k} \leq m$, $1 \leq j < \hat{j} < l \leq n$. The S-polynomial is then of the form

$$\begin{aligned} S(f_{ijkl}, f_{i\hat{j}\hat{k}l}) &= x_{\hat{k}\hat{j}} f_{ijkl}(\mathbf{x}) - x_{kj} f_{i\hat{j}\hat{k}l}(\mathbf{x}) = -x_{ij}x_{kl}x_{\hat{k}\hat{j}} + x_{i\hat{j}}x_{\hat{k}l}x_{kj} \\ &= x_{\hat{k}l} f_{ijk\hat{j}}(\mathbf{x}) - x_{ij} f_{k\hat{j}\hat{k}l}(\mathbf{x}) \in J_{M_{mn}} \end{aligned}$$

so that $S(f_{ijkl}, f_{i\hat{j}\hat{k}l}) \to_{\mathcal{G}_{M_{mn}}} 0$. The remaining cases are treated with similar arguments.

In order to show that $\mathcal{G}_{M_{mn}}$ is a reduced Gröbner basis (see Definition 8), we first notice that $\mathrm{LC}(f) = 1$ for all $f \in \mathcal{G}_{M_{mn}}$. In addition, the leading monomial of $f \in \mathcal{G}_{M_{mn}}$ is always of degree two and there are no two different polynomials $f_i, f_j \in \mathcal{G}_{M_{mn}}$ such that $\mathrm{LM}(f_i) = \mathrm{LM}(f_j)$. Therefore, $\mathcal{G}_{M_{mn}}$ is the reduced Gröbner basis of the ideal $J_{M_{mn}}$ with respect to the grevlex order. □

The Gröbner basis $\mathcal{G}_{M_{mn}}$ of $J_{M_{mn}} = \langle \mathcal{G}_{M_{mn}} \rangle$ yields the θ-basis of $\mathbb{R}[\mathbf{x}]/J_{M_{mn}}$. For the sake of simplicity, we only provide its elements up to degree two,

$$\mathcal{B}_1 = \{1 + J_{M_{mn}}, x_{11} + J_{M_{mn}}, x_{12} + J_{M_{mn}}, \ldots, x_{mn} + J_{M_{mn}}\}$$
$$\mathcal{B}_2 = \mathcal{B}_1 \cup \{x_{ij}x_{kl} + J_{M_{mn}} : (i,j,k,l) \in \mathcal{S}_{\mathcal{B}_2}\},$$

where $\mathcal{S}_{\mathcal{B}_2} = \{(i,j,k,l) : 1 \le i \le k \le m, 1 \le j \le l \le n\} \setminus (1,1,1,1)$. Given the θ-basis, the theta body $\mathrm{TH}_k(J_{M_{mn}})$ is well-defined. We formally introduce an associated norm next.

Definition 5. The matrix θ_k*-norm*, denoted by $\|\cdot\|_{\theta_k}$, is the norm induced by the k-theta body $\mathrm{TH}_k(J_{M_{mn}})$, i.e.,

$$\|\mathbf{X}\|_{\theta_k} = \inf\{r : \mathbf{X} \in r\,\mathrm{TH}_k(J_{M_{mn}})\}.$$

The θ_k-norm can be computed with the help of Theorem 1, i.e., as

$$\|\mathbf{X}\|_{\theta_k} = \min t \quad \text{subject to } \mathbf{X} \in tQ_{\mathcal{B}_k}(J_{M_{mn}}).$$

Given the moment matrix $\mathbf{M}_{\mathcal{B}_k}[\mathbf{y}]$ associated with J, this minimization program is equivalent to the semidefinite program

$$\min_{t\in\mathbb{R}, \mathbf{y}\in\mathbb{R}^{\mathcal{B}_k}} t \quad \text{subject to} \quad \mathbf{M}_{\mathcal{B}_k}[\mathbf{y}] \succcurlyeq 0, y_0 = t, \mathbf{y}_{\mathcal{B}_1} = \mathbf{X}. \tag{9}$$

The last constraint might require some explanation. The vector $\mathbf{y}_{\mathcal{B}_1}$ denotes the restriction of $\mathbf{y}$ to the indices in $\mathcal{B}_1$, where the latter can be identified with the set $[m]\times[n]$ indexing the matrix entries. Therefore, $\mathbf{y}_{\mathcal{B}_1} = \mathbf{X}$ means componentwise $y_{x_{11}+J} = X_{11}, y_{x_{12}+J} = X_{12}, \ldots, y_{x_{mn}+J} = X_{mn}$. For the purpose of illustration, we focus on the θ_1-norm in $\mathbb{R}^{2\times 2}$ in Section 3.1 below, and provide a step-by-step procedure for building the corresponding semidefinite program in (9).

Notice that the number of elements in $\mathcal{B}_1$ is $mn+1$, and in $\mathcal{B}_2\backslash\mathcal{B}_1$ is $\frac{m\cdot(m+1)}{2}\cdot\frac{n\cdot(n+1)}{2} - 1 \sim \frac{(mn)^2}{2}$, i.e., the number of elements of the θ-basis restricted to the degree 2 scales polynomially in the total number of matrix entries mn. Therefore, the computational complexity of the SDP in (9) is polynomial in mn.

We will show next that the theta body $\mathrm{TH}_1(J)$ and hence, all $\mathrm{TH}_k(J)$ for $k \in \mathbb{N}$, coincide with the nuclear norm ball. To this end, the following lemma provides expressions for the boundary of the matrix nuclear unit norm ball.

Lemma 3. *Let $\mathcal{O}_c$ ($\mathcal{O}_r$) denote the set of all matrices $\mathbf{M} \in \mathbb{R}^{n\times m}$ with orthonormal columns (rows), i.e., $\mathcal{O}_c = \{\mathbf{M} \in \mathbb{R}^{n\times m} : \mathbf{M}^T\mathbf{M} = \mathbf{I}_m\}$ and $\mathcal{O}_r = \{\mathbf{M} \in \mathbb{R}^{n\times m} : \mathbf{M}\mathbf{M}^T = \mathbf{I}_n\}$. Then*

$$\{\mathbf{X} \in \mathbb{R}^{m\times n} : \|\mathbf{X}\|_* \le 1\} = \{\mathbf{X} \in \mathbb{R}^{m\times n} : \mathrm{tr}(\mathbf{M}\mathbf{X}) \le 1, \text{ for all } \mathbf{M} \in \mathcal{O}_c \cup \mathcal{O}_r\}. \tag{10}$$

Remark 1. Notice that $\mathcal{O}_c = \emptyset$ for $m > n$ and $\mathcal{O}_r = \emptyset$ for $m < n$.

Proof. If suffices to treat the case $m \le n$ because $\|\mathbf{X}\|_* = \|\mathbf{X}^T\|_*$ for all matrices $\mathbf{X}$, and $\mathbf{M} \in \mathcal{O}_r$ if and only if $\mathbf{M}^T \in \mathcal{O}_c$. Let $\mathbf{X} \in \mathbb{R}^{m\times n}$ such that $\|\mathbf{X}\|_* \le 1$ and let $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ be its singular value decomposition. For $\mathbf{M} \in \mathcal{O}_c$, the spectral norm satisfies $\|\mathbf{M}\| \le 1$ and therefore, using that the nuclear norm is the dual of the spectral norm, see e.g. [1, p. 96],

$$\mathrm{tr}(\mathbf{M}\mathbf{X}) \le \|\mathbf{M}\| \cdot \|\mathbf{X}\|_* \le \|\mathbf{X}\|_* \le 1.$$

For the converse, let $\mathbf{X} \in \mathbb{R}^{m\times n}$ be such that $\mathrm{tr}(\mathbf{M}\mathbf{X}) \le 1$, for all $\mathbf{M} \in \mathcal{O}_c$. Let $\mathbf{X} = \mathbf{U}\overline{\mathbf{\Sigma}}\overline{\mathbf{V}}^T$ denote its reduced singular value decomposition, i.e., $\mathbf{U}, \overline{\mathbf{\Sigma}} \in \mathbb{R}^{m\times m}$ and $\overline{\mathbf{V}} \in \mathbb{R}^{n\times m}$ with $\mathbf{U}^T\mathbf{U} = \mathbf{U}\mathbf{U}^T = \overline{\mathbf{V}}^T\overline{\mathbf{V}} = \mathbf{I}_m$. Since $\mathbf{M} := \overline{\mathbf{V}}\mathbf{U}^T \in \mathcal{O}_c$, it follows that

$$1 \ge \mathrm{tr}(\mathbf{M}\mathbf{X}) = \mathrm{tr}(\overline{\mathbf{V}}\mathbf{U}^T\mathbf{U}\mathbf{\Sigma}\overline{\mathbf{V}}^T) = \mathrm{tr}(\mathbf{\Sigma}) = \|\mathbf{X}\|_*.$$

This completes the proof. □

Next, using Lemma 3, we show that the theta body $\mathrm{TH}_1(J)$ equals the nuclear norm ball. This result is related to Theorem 4.4 in [23].

Theorem 2. *The polynomial ideal $J_{M_{mn}}$ defined in (8) is TH_1-exact, i.e.,*

$$\mathrm{TH}_1\left(J_{M_{mn}}\right) = \operatorname{conv}\left(\mathbf{x} : g(\mathbf{x}) = 0, f_{ijkl}(\mathbf{x}) = 0 \text{ for all } i < k, j < l\right).$$

In other words,

$$\left\{\mathbf{X} \in \mathbb{R}^{m\times n} : \mathbf{X} \in \mathrm{TH}_1\left(J_{M_{mn}}\right)\right\} = \left\{\mathbf{X} \in \mathbb{R}^{m\times n} : \|\mathbf{X}\|_* \leq 1\right\}.$$

Proof. By definition of $\mathrm{TH}_1(J)$, it is enough to show that the boundary of the unit nuclear norm can be written as 1-sos mod $J_{M_{mn}}$, which by Lemma 3 means that the polynomial $1 - \sum_{i=1}^m \sum_{j=1}^n x_{ij} M_{ji}$ is 1-sos mod $J_{M_{mn}}$ for all $\mathbf{M} \in \mathcal{O}_c \cup \mathcal{O}_r$. We start by fixing $\mathbf{M} = \begin{pmatrix}\mathbf{I}_m \\ \mathbf{0}\end{pmatrix}$ in case $m \leq n$ and $\mathbf{M} = \begin{pmatrix}\mathbf{I}_n & \mathbf{0}\end{pmatrix}$ in case $m > n$, where $\mathbf{I}_k \in \mathbb{R}^{k\times k}$ is the identity matrix. For this choice of $\mathbf{M}$, we need to show that $1 - \sum_{i=1}^{\ell} x_{ii}$ is 1-sos mod $J_{M_{mn}}$, where $\ell = \min\{m,n\}$. Note that

$$\begin{aligned}1 - \sum_{i=1}^{\ell} x_{ii} =& \frac{1}{2}\left[\left(1 - \sum_{i=1}^{\ell} x_{ii}\right)^2 + \left(1 - \sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij}^2\right) + \sum_{i<j\leq \ell}(x_{ij} - x_{ji})^2\right.\\ &\left. -2\sum_{i<j\leq\ell}(x_{ii}x_{jj} - x_{ij}x_{ji}) + \sum_{i=1}^{m}\sum_{j=m+1}^{n} x_{ij}^2 + \sum_{i=n+1}^{m}\sum_{j=1}^{n} x_{ij}^2\right],\end{aligned}$$

since

$$\left(1 - \sum_{i=1}^{\ell} x_{ii}\right)^2 = 1 - 2\sum_{i=1}^{\ell} x_{ii} + \sum_{i=1}^{\ell}\sum_{j=1}^{\ell} x_{ii}x_{jj} = 1 - 2\sum_{i=1}^{\ell} x_{ii} + 2\sum_{i<j\leq\ell} x_{ii}x_{jj} + \sum_{i=1}^{\ell} x_{ii}^2,$$

$$1 - \sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij}^2 + \sum_{i=1}^{m}\sum_{j=m+1}^{n} x_{ij}^2 + \sum_{i=n+1}^{m}\sum_{j=1}^{n} x_{ij}^2 = 1 - \sum_{i=1}^{\ell}\sum_{j=1}^{\ell} x_{ij}^2 = 1 - \sum_{i<j\leq\ell}\left(x_{ij}^2 + x_{ji}^2\right) - \sum_{i=1}^{\ell} x_{ii}^2,$$

and

$$\begin{aligned}\sum_{i<j\leq\ell}(x_{ij} - x_{ji})^2 - 2\sum_{i<j\leq\ell}(x_{ii}x_{jj} - x_{ij}x_{ji}) &= \sum_{i<j\leq\ell}\left(x_{ij}^2 + x_{ji}^2 - 2x_{ij}x_{ji} - 2x_{ii}x_{jj} + 2x_{ij}x_{ji}\right)\\ &= \sum_{i<j\leq\ell}\left(x_{ij}^2 + x_{ji}^2\right) - 2\sum_{i<j\leq\ell} x_{ii}x_{jj}.\end{aligned}$$

Therefore, $1 - \sum_{i=1}^{\ell} x_{ii}$ is 1-sos mod $J_{M_{mn}}$, since the polynomials $1 - \sum_{i=1}^{\ell} x_{ii}$, $x_{ij} - x_{ji}$, x_{ij}, and x_{ji} are linear and the polynomials $1 - \sum_{i=1}^m \sum_{j=1}^n x_{ij}^2$ and $2\left(x_{ii}x_{jj} - x_{ij}x_{ji}\right)$ are contained in the ideal, for all $i < j \leq \ell$.

Next, we define transformed variables

$$x'_{ij} = \begin{cases}\sum_{k=1}^m M_{ik}x_{kj} & \text{if } m \leq n,\\ \sum_{k=1}^n x_{ik}M_{kj} & \text{if } m > n.\end{cases}$$

Since x'_{ij} is a linear combination of $\{x_{kj}\}_{k=1}^m \cup \{x_{ik}\}_{k=1}^n$, for every $i \in [m]$ and $j \in [n]$, linearity of the polynomials $1 - \sum_{i=1}^{\ell} x'_{ii}$, $x'_{ij} - x'_{ji}$, x'_{ij}, and x'_{ji} is preserved, for all $i < j$. It remains to show that the ideal is invariant under this transformation. For the polynomial $1 - \sum_{i=1}^m \sum_{j=1}^n {x'_{ij}}^2$ this is clear since $\mathbf{M} \in \mathbb{R}^{n\times m}$ has unitary columns in case when $m \leq n$ and unitary rows in case $m \geq n$. In the case of $m \leq n$ the polynomial $x'_{ii}x'_{jj} - x'_{ij}x'_{ji}$ is contained in the ideal J since

$$x'_{ii}x'_{jj} - x'_{ij}x'_{ji} = \sum_{k=1}^{m}\sum_{l=1}^{m} M_{ik}M_{jl}\left(x_{ki}x_{lj} - x_{kj}x_{li}\right)$$

and the polynomials $x_{ki}x_{lj} - x_{kj}x_{li}$ are contained in J for all $i < j \leq m$. Similarly, in case $m \geq n$ the polynomial $x'_{ii}x'_{jj} - x'_{ij}x'_{ji}$ is in the ideal since

$$x'_{ii}x'_{jj} - x'_{ij}x'_{ji} = \sum_{k=1}^{n}\sum_{l=1}^{n} M_{ki}M_{lj}\left(x_{ik}x_{jl} - x_{il}x_{jk}\right)$$

and polynomials $x_{ik}x_{jl} - x_{il}x_{jk}$ are in the ideal, for all $i < j \leq n$. □

The following corollary is a direct consequence of Theorem 2 and the nestedness property (4) of theta bodies.

Corollary 1. *The matrix θ_1-norm coincides with the matrix nuclear norm, i.e.,*

$$\|\mathbf{X}\|_* = \|\mathbf{X}\|_{\theta_1}, \quad \textit{for all } \mathbf{X} \in \mathbb{R}^{m \times n}.$$

Moreover,

$$\mathrm{TH}_1\left(J_{M_{mn}}\right) = \mathrm{TH}_2\left(J_{M_{mn}}\right) = \cdots = \mathrm{conv}\left(\nu_{\mathbb{R}}\left(J_{M_{mn}}\right)\right).$$

Remark 2. The ideal (8) is not the only choice that satisfies (5). For example, in [10] the following polynomial ideal was suggested

$$J = \left\langle \left\{x_{ij} - u_i v_j\right\}_{i\in[m], j\in[n]}, \sum_{i=1}^{m} u_i^2 - 1, \sum_{j=1}^{n} v_j^2 - 1 \right\rangle \tag{11}$$

in $\mathbb{R}[\mathbf{x}, \mathbf{u}, \mathbf{v}] = \mathbb{R}[x_{11}, \ldots, x_{mn}, u_1, \ldots, u_m, v_1, \ldots, u_n]$. Some tedious computations reveal the reduced Gröbner basis $\mathcal{G}$ of the ideal J with respect to the grevlex (and grlex) ordering,

$$\begin{aligned}
\mathcal{G} = &\left\{g_1^{ij} = x_{ij} - u_i v_j : i \in [m], j \in [n]\right\} \bigcup \left\{g_2 = \sum_{i=1}^{m} u_i^2 - 1, g_3 = \sum_{j=1}^{n} v_j^2 - 1\right\} \\
&\bigcup \left\{g_4^{i,j,k} = x_{ij}u_k - x_{kj}u_i : 1 \leq i < k \leq m, j \in [n]\right\} \bigcup \left\{g_5^{j} = \sum_{i=1}^{m} x_{ij}u_i - v_j : j \in [n]\right\} \\
&\bigcup \left\{g_6^{i,j,k} = x_{ij}v_k - x_{ik}v_j : i \in [m], 1 \leq j < k \leq n\right\} \bigcup \left\{g_7^{i} = \sum_{j=1}^{n} x_{ij}v_j - u_i : i \in [m]\right\} \\
&\bigcup \left\{g_8^{i,j} = \sum_{k=1}^{n} x_{ik}x_{jk} - u_i u_j : 1 \leq i < j \leq m\right\} \bigcup \left\{g_9^{i,j} = \sum_{k=1}^{m} x_{ki}x_{kj} - v_i v_j : 1 \leq i < j \leq n\right\} \\
&\bigcup \left\{g_{10}^{i} = \sum_{j=1}^{n} x_{ij}^2 - u_i^2 : 2 \leq i \leq m\right\} \bigcup \left\{g_{11}^{j} = \sum_{i=1}^{m} x_{ij}^2 - v_j^2 : 2 \leq j \leq n\right\} \\
&\bigcup \left\{g_{12}^{i,j,k,l} = x_{ij}x_{kl} - x_{il}x_{kj} : 1 \leq i < k \leq m, 1 \leq j < l \leq n\right\} \\
&\bigcup \left\{g_{13} = x_{11}^2 - \sum_{i=2}^{m}\sum_{j=2}^{n} x_{ij}^2 + \sum_{i=2}^{m} u_i^2 + \sum_{j=2}^{n} v_j^2 - 1\right\}.
\end{aligned} \tag{12}$$

Obviously, this Gröbner basis is much more complicated than the one of the ideal $J_{M_{mn}}$ introduced above. Therefore, computations (both theoretical and numerical) with this alternative ideal seem to be more demanding. In any case, the variables $\{u_i\}_{i=1}^m$ and $\{v_j\}_{j=1}^n$ are only auxiliary ones, so one would like to eliminate these from the above Gröbner basis. By doing so, one obtains the Gröbner basis $\mathcal{G}_{M_{mn}}$ defined in (8). Notice that $\sum_{i=1}^m \sum_{j=1}^n x_{ij}^2 - 1 = g_{13} + \sum_{i=2}^m g_{10}^i + \sum_{j=2}^n g_{11}^j$ together with $\{g_{12}^{i,j,k,l}\}$ form the basis $\mathcal{G}_{M_{mn}}$.

1	x_{11}	x_{12}	x_{21}	x_{22}	$x_{11}x_{12}$	$x_{11}x_{21}$	$x_{11}x_{22}$	x_{12}^2	$x_{12}x_{22}$	x_{21}^2	$x_{21}x_{22}$	x_{22}^2
y_0	x_{11}	x_{12}	x_{21}	x_{22}	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8

TABLE 2. Linearization of the elements of $\mathcal{B}_2$ for matrix 2×2 case. The polynomial f in the first row refers to the element $f + J \in \mathcal{B}_2$.

3.1. **The θ_1-norm in $\mathbb{R}^{2\times 2}$.** For the sake of illustration, we consider the specific example of 2×2 matrices and provide the corresponding semidefinite program for the computation of the θ_1-norm explicitly. Let us denote the corresponding polynomial ideal in $\mathbb{R}[\mathbf{x}] = \mathbb{R}[x_{11}, x_{12}, x_{21}, x_{22}]$ simply by

$$J = J_{M_{22}} = \left\langle x_{12}x_{21} - x_{11}x_{22}, x_{11}^2 + x_{12}^2 + x_{21}^2 + x_{22}^2 - 1 \right\rangle \tag{13}$$

The associated algebraic variety is of the form

$$\nu_{\mathbb{R}}(J) = \left\{ \mathbf{x} : x_{12}x_{21} = x_{11}x_{22}, x_{11}^2 + x_{12}^2 + x_{21}^2 + x_{22}^2 = 1 \right\}$$

and corresponds to the set of rank one matrices with $\|\mathbf{X}\|_F = 1$. Its convex hull consists of matrices $\mathbf{X} \in \mathbb{R}^{2\times 2}$ with $\|\mathbf{X}\|_* \leq 1$. According to Lemma 2, the Gröbner basis $\mathcal{G}$ of J with respect to the grevlex order is

$$\mathcal{G} = \left\{ g_1 = x_{12}x_{21} - x_{11}x_{22},\ g_2 = x_{11}^2 + x_{12}^2 + x_{21}^2 + x_{22}^2 - 1 \right\}$$

with the corresponding θ-basis $\mathcal{B}$ of $\mathbb{R}[\mathbf{x}]/J$ restricted to the degree two given as

$$\mathcal{B}_1 = \{1 + J, x_{11} + J, x_{12} + J, x_{21} + J, x_{22} + J\}$$

$$\mathcal{B}_2 = \mathcal{B}_1 \cup \left\{ x_{11}x_{12} + J, x_{11}x_{21} + J, x_{11}x_{22} + J, x_{12}^2 + J, x_{12}x_{22} + J, x_{21}^2 + J, x_{21}x_{22} + J, x_{22}^2 + J \right\}.$$

The set $\mathcal{B}_2$ consists of all monomials of degree at most two which are not divisible by a leading term of any of the polynomials inside the Gröbner basis $\mathcal{G}$. For example, $x_{11}x_{12} + J$ is an element of the theta basis $\mathcal{B}$, but $x_{11}^2 + J$ is not since x_{11}^2 is divisible by $\mathrm{LT}(g_2)$.

Linearizing the elements of $\mathcal{B}_2$ results in Table 2, where the monomials f in the first row stand for an element $f + J \in \mathcal{B}_2$. Therefore, $[\mathbf{x}]_{\mathcal{B}_1} = (1, x_{11}, x_{12}, x_{21}, x_{22})^T$ and the following combinatorial moment matrix $\mathbf{M}_{\mathcal{B}_1}(\mathbf{x}, \mathbf{y})$, see Definition 4, is given as

$$\mathbf{M}_{\mathcal{B}_1}(\mathbf{x}, \mathbf{y}) = \begin{bmatrix} y_0 & x_{11} & x_{12} & x_{21} & x_{22} \\ x_{11} & -y_4 - y_6 - y_8 + y_0 & y_1 & y_2 & y_3 \\ x_{12} & y_1 & y_4 & y_3 & y_5 \\ x_{21} & y_2 & y_3 & y_6 & y_7 \\ x_{22} & y_3 & y_5 & y_7 & y_8 \end{bmatrix}.$$

For instance, the entry $(2,2)$ of $[\mathbf{x}]_{\mathcal{B}_1}[\mathbf{x}]_{\mathcal{B}_1}^T$ is of the form $x_{11}^2 + J = -x_{12}^2 - x_{21}^2 - x_{22}^2 + 1 + J$, where we exploit the second property in Definition 3 and the fact that $g_2 \in J$. Replacing $x_{12}^2 + J$ by y_4, etc. as in Table 2, yields the stated expression for $\mathbf{M}_{\mathcal{B}_1}(\mathbf{x}, \mathbf{y})_{2,2}$.

By Theorem 1, the first theta body $\mathrm{TH}_1(J)$ is the closure of

$$\mathbf{Q}_{\mathcal{B}_1}(J) = \pi_{\mathbf{x}} \left\{ (\mathbf{x}, \mathbf{y}) \in \mathbb{R}^{\mathcal{B}_2} : \mathbf{M}_{\mathcal{B}_1}(\mathbf{x}, \mathbf{y}) \succeq 0,\ y_0 = 1 \right\},$$

where $\pi_{\mathbf{x}}$ represents the projection onto the variables $\mathbf{x}$, i.e., the projection onto x_{11}, x_{12}, x_{21}, x_{22}. Furthermore, θ_1-norm of a matrix $\mathbf{X} \in \mathbb{R}^{2\times 2}$ induced by the $\mathrm{TH}_1(J)$ and denoted as $\|\cdot\|_{\theta_1}$ can be computed as

$$\|\mathbf{X}\|_{\theta_1} = \inf t \text{ s.t. } \mathbf{X} \in t\mathbf{Q}_{\mathcal{B}_1}(J) \tag{14}$$

which is equivalent to

$$\inf_{t \in \mathbb{R}, \mathbf{y} \in \mathbb{R}^8} t \quad \text{s.t.} \quad \mathbf{M} = \begin{bmatrix} t & X_{11} & X_{12} & X_{21} & X_{22} \\ X_{11} & -y_4 - y_6 - y_8 + t & y_1 & y_2 & y_3 \\ X_{12} & y_1 & y_4 & y_3 & y_5 \\ X_{21} & y_2 & y_3 & y_6 & y_7 \\ X_{22} & y_3 & y_5 & y_7 & y_8 \end{bmatrix} \succeq 0. \tag{15}$$

Notice that $\text{trace}(\mathbf{M}) = 2t$. By Theorem 2, the above program is equivalent to the standard semidefinite program for computing the nuclear norm of a given matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$

$$\min_{\mathbf{W},\mathbf{Z}} \frac{1}{2} \left(\text{trace}\left(\mathbf{W}\right) + \text{trace}\left(\mathbf{Z}\right)\right) \quad \text{s.t.} \quad \begin{bmatrix} W_{11} & W_{12} & X_{11} & X_{12} \\ W_{12} & W_{22} & X_{21} & X_{22} \\ X_{11} & X_{21} & Z_{11} & Z_{12} \\ X_{22} & X_{22} & Z_{12} & Z_{22} \end{bmatrix} \succeq 0.$$

Notice that the matrix $\mathbf{M}$ in (15) can be written as the following sum

$$\mathbf{M} = t \cdot \mathbf{M}_0 + \sum_{i=1}^{2} \sum_{j=1}^{2} X_{ij} \mathbf{M}_{ij} + \sum_{k=1}^{8} y_k \mathbf{M}_k,$$

where

$$\mathbf{M}_0 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{M}_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \mathbf{M}_2 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{M}_3 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix},$$

$$\mathbf{M}_4 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{M}_5 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}, \mathbf{M}_6 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{M}_7 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix},$$

$$\mathbf{M}_8 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \mathbf{M}_{11} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \mathbf{M}_{12} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{M}_{21} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$\mathbf{M}_{22} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

4. The tensor θ_k-norm

Let us now turn to the tensor case and study the hierarchical closed convex relaxations of the tensor unit nuclear norm ball defined via theta bodies. Since in the matrix case all θ_k-norms are equal to the matrix nuclear norm, their generalization to the tensor case may all be viewed as natural generalizations of the nuclear norm. We focus mostly on the θ_1-norm whose unit norm ball is the largest in a hierarchical sequence of relaxations. Unlike in the matrix case, the θ_1-norm defines a new tensor norm, that up to the best of our knowledge has not yet been studied before.

The polynomial ideal will be generated by the minors of order two of the unfoldings – and matricizations in the case $d \geq 4$ – of the tensors, where each variable corresponds to one entry in the tensor. As we will see, a tensor is of rank one if and only if all order two minors of the unfoldings (matricizations) vanish. While the order three case requires to consider all three unfoldings, there are several possibilities for the order d case when $d \geq 4$. In fact, a d-th order tensor is of rank one if all minors of all unfoldings vanish so that it may be enough to consider only the unfoldings. However, one may as well consider the ideal generated by all minors of *all matricizations* or one may consider a subset of matricizations including all unfoldings. Indeed, any *tensor format* – and thereby any notion of tensor rank – corresponds to a set of matricizations and in this way, one may associate a θ_k-norm to a certain tensor format. We refer to e.g. [28, 45] for some background on various tensor formats. We will mainly concentrate on the case that *all* matricizations are taken into account for defining the ideal. Only for the case $d = 4$, we will briefly discuss the case, that the ideal is generated only by the minors corresponding to the four unfoldings.

Below, we consider first the special case of third order tensors and continue then with fourth order tensors. In Subsection 4.3 we will treat the general dth order case.

4.1. **Third order tensors.** As described above, we will consider the minors of order two of all unfoldings of a third order tensor. Our notation requires the following sets of subscripts

$$\mathcal{S}_0 = \left\{ijk\hat{i}\hat{j}\hat{k} : 1 \leq i < \hat{i} \leq n_1,\ 1 \leq j < \hat{j} \leq n_2,\ 1 \leq k < \hat{k} \leq n_3\right\}, \tag{16}$$

$$\mathcal{S}_1 = \left\{ijk\hat{i}\hat{j}\hat{k} : 1 \leq i < \hat{i} \leq n_1,\ 1 \leq j < \hat{j} \leq n_2,\ 1 \leq k \leq \hat{k} \leq n_3\right\}, \tag{17}$$

$$\mathcal{S}_2 = \left\{ijk\hat{i}\hat{j}\hat{k} : 1 \leq i \leq \hat{i} \leq n_1,\ 1 \leq j < \hat{j} \leq n_2,\ 1 \leq k < \hat{k} \leq n_3\right\}, \tag{18}$$

$$\mathcal{S}_3 = \left\{ijk\hat{i}\hat{j}\hat{k} : 1 \leq i < \hat{i} \leq n_1,\ 1 \leq j \leq \hat{j} \leq n_2,\ 1 \leq k < \hat{k} \leq n_3\right\}. \tag{19}$$

The following polynomials in $\mathbb{R}[\mathbf{x}] = \mathbb{R}[x_{111}, x_{112}, \ldots, x_{n_1 n_2 n_3}]$ correspond to all order two minors,

$$f_1^{ijk\hat{i}\hat{j}\hat{k}}(\mathbf{x}) = -x_{ijk}x_{\hat{i}\hat{j}\hat{k}} + x_{ij\hat{k}}x_{\hat{i}jk}, \quad ijk\hat{i}\hat{j}\hat{k} \in \mathcal{S}_1, \tag{20}$$

$$f_2^{ijk\hat{i}\hat{j}\hat{k}}(\mathbf{x}) = -x_{ijk}x_{\hat{i}\hat{j}\hat{k}} + x_{i\hat{j}\hat{k}}x_{\hat{i}jk}, \quad ijk\hat{i}\hat{j}\hat{k} \in \mathcal{S}_2, \tag{21}$$

$$f_3^{ijk\hat{i}\hat{j}\hat{k}}(\mathbf{x}) = -x_{ijk}x_{\hat{i}\hat{j}\hat{k}} + x_{i\hat{j}k}x_{\hat{i}j\hat{k}}, \quad ijk\hat{i}\hat{j}\hat{k} \in \mathcal{S}_3, \tag{22}$$

$$g_3(\mathbf{x}) = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3} x_{ijk}^2 - 1. \tag{23}$$

Lemma 4. *A tensor $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is a rank one, Frobenius norm one tensor if and only if*

$$g_3(\mathbf{X}) = 0 \ \text{ and } \ f_\ell^{ijk\hat{i}\hat{j}\hat{k}}(\mathbf{X}) = 0 \quad \text{for all} \quad ijk\hat{i}\hat{j}\hat{k} \in \mathcal{S}_\ell,\ \ell \in [3]. \tag{24}$$

Proof. Sufficiency of (24) follows directly from the definition of the rank one Frobenius norm one tensors. For necessity, the first step is to show that mode-1 fibers (columns) span one dimensional space in $\mathbb{R}^{n_1}$. To this end, we note that, for $j \leq \hat{j}$ and $k \leq \hat{k}$, the fibers $\mathbf{X}_{\cdot jk}$ and $\mathbf{X}_{\cdot \hat{j}\hat{k}}$ satisfy

$$-X_{n_1\hat{j}\hat{k}}\begin{bmatrix} X_{1jk} \\ X_{2jk} \\ \vdots \\ X_{n_1jk}\end{bmatrix} + X_{n_1jk}\begin{bmatrix} X_{1\hat{j}\hat{k}} \\ X_{2\hat{j}\hat{k}} \\ \vdots \\ X_{n_1\hat{j}\hat{k}}\end{bmatrix} \overset{(20),(22)}{=} \begin{bmatrix} -X_{1\hat{j}\hat{k}}X_{n_1jk} + X_{1\hat{j}\hat{k}}X_{n_1jk} \\ -X_{2\hat{j}\hat{k}}X_{n_1jk} + X_{2\hat{j}\hat{k}}X_{n_1jk} \\ \vdots \\ -X_{n_1\hat{j}\hat{k}}X_{n_1jk} + X_{n_1\hat{j}k}X_{n_1j\hat{k}}\end{bmatrix} = \mathbf{0},$$

where we used (20) for $j < \hat{j}$ and $k \leq \hat{k}$ and (22) for $j = \hat{j}$ and $k < \hat{k}$. From (23) and (24) it follows that the tensor $\mathbf{X}$ is normalized.

Using similar arguments, one argues that mode-2 fibers (rows) and mode-3 fibers span one dimensional spaces in $\mathbb{R}^{n_2}$ and $\mathbb{R}^{n_3}$, respectively. This completes the proof. □

A third order tensor $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is a rank-one tensor if and only if all three unfoldings $\mathbf{X}^{\{1\}} \in \mathbb{R}^{n_1 \times n_2 n_3}$, $\mathbf{X}^{\{2\}} \in \mathbb{R}^{n_2 \times n_1 n_3}$, and $\mathbf{X}^{\{3\}} \in \mathbb{R}^{n_3 \times n_1 n_2}$ are rank-one matrices. Notice that $f_\ell^{ijk\hat{i}\hat{j}\hat{k}}(\mathbf{X}) = 0$ for all $ijk\hat{i}\hat{j}\hat{k} \in \mathcal{S}_\ell$ is equivalent to the statement that the ℓ-th unfolding $\mathbf{X}^{\{\ell\}}$ is a rank one matrix, i.e., that all its minors of order two vanish, for all $\ell \in [3]$. In order to define relaxations of the tensor nuclear norm ball we introduce the polynomial ideal $J_3 \subset \mathbb{R}[\mathbf{x}] = \mathbb{R}[x_{111}, x_{112}, \ldots, x_{n_1 n_2 n_3}]$ as the one generated by

$$\mathcal{G}_3 = \left\{f_\ell^{ijk\hat{i}\hat{j}\hat{k}}(\mathbf{x}) : ijk\hat{i}\hat{j}\hat{k} \in \mathcal{S}_\ell, \ell \in [3]\right\} \cup \{g_3(\mathbf{x})\}, \tag{25}$$

i.e., $J_3 = \langle \mathcal{G}_3 \rangle$. Its algebraic variety equals the set of rank one order three tensors with unit Frobenius norm and its convex hull coincides with the tensor nuclear norm ball. The next result provides the Gröbner basis of J_3.

Theorem 3. *The basis $\mathcal{G}_3$ defined in (25) forms the reduced Gröbner basis of the ideal $J_3 = \langle \mathcal{G}_3 \rangle$ with respect to the grevlex order.*

	$\mathbf{X} \in \mathbb{R}^{2\times 2\times 2}$	$\|\mathbf{X}^{\{1\}}\|_*$	$\|\mathbf{X}^{\{2\}}\|_*$	$\|\mathbf{X}^{\{3\}}\|_*$	$\|\mathbf{X}\|_{\theta_1}$
1	$\left[\begin{array}{cc\|cc} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right]$	2	2	2	2
2	$\left[\begin{array}{cc\|cc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{array}\right]$	2	2	$\sqrt{2}$	2
3	$\left[\begin{array}{cc\|cc} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right]$	2	$\sqrt{2}$	2	2
4	$\left[\begin{array}{cc\|cc} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right]$	$\sqrt{2}$	2	2	2
5	$\left[\begin{array}{cc\|cc} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{array}\right]$	$\sqrt{2}+1$	$\sqrt{2}+1$	$\sqrt{2}+1$	3

TABLE 3. Matrix nuclear norms of unfoldings and θ_1-norm of tensors $\mathbf{X} \in \mathbb{R}^{2\times 2\times 2}$, which are represented in the second column as $\mathbf{X} = [\mathbf{X}(:,:,1) \,|\, \mathbf{X}(:,:,2)]$. The third, fourth and fifth column represent the nuclear norms of the first, second and the third unfolding of a tensor $\mathbf{X}$, respectively. The last column contains the numerically computed θ_1-norm.

Proof. Similarly as in the proof of Theorem 2 we need to show that $S(p,q) \to_{\mathcal{G}_3} 0$ for all relatively prime polynomials $p, q \in \mathcal{G}_3$. The leading monomials with respect to the grevlex ordering are given by $\mathrm{LM}(g_3) = x_{111}^2$ and

$$\mathrm{LM}(f_1^{ijk\hat{i}\hat{j}\hat{k}}) = x_{\hat{i}\hat{j}\hat{k}} x_{\hat{i}jk}, \quad ijk\hat{i}\hat{j}\hat{k} \in \mathcal{S}_1,$$
$$\mathrm{LM}(f_2^{ijk\hat{i}\hat{j}\hat{k}}) = x_{\hat{i}\hat{j}k} x_{ij\hat{k}}, \quad ijk\hat{i}\hat{j}\hat{k} \in \mathcal{S}_2,$$
$$\mathrm{LM}(f_3^{ijk\hat{i}\hat{j}\hat{k}}) = x_{i\hat{j}k} x_{\hat{i}j\hat{k}}, \quad ijk\hat{i}\hat{j}\hat{k} \in \mathcal{S}_3.$$

The polynomial g_3 is relatively prime with every other polynomial in the basis $\mathcal{G}_3$. First we consider two distinct polynomials $f, g \in \{f_3^{ijk\hat{i}\hat{j}\hat{k}} : ijk\hat{i}\hat{j}\hat{k} \in \mathcal{S}_3\}$. So let $f = f_3^{ijk\hat{i}\hat{j}\hat{k}}$ and $g = f_3^{ijk\hat{i}j\hat{k}}$ for $ijk\hat{i}\hat{j}\hat{k} \in \mathcal{S}_0$, i.e.,

$$f = -x_{ijk}x_{\hat{i}\hat{j}\hat{k}} + x_{ij\hat{k}}x_{\hat{i}\hat{j}k}, \qquad g = -x_{ijk}x_{\hat{i}j\hat{k}} + x_{ij\hat{k}}x_{\hat{i}jk}.$$

Then

$$S(f,g) = x_{ijk}\left(-x_{\hat{i}\hat{j}\hat{k}}x_{\hat{i}jk} + x_{\hat{i}j\hat{k}}x_{\hat{i}\hat{j}k}\right) = x_{ijk} f_2^{\hat{i}jk\hat{i}\hat{j}\hat{k}} \to_{\mathcal{G}_3} 0.$$

Next we show that $S(f,g) \in J_3$, for $f \in \{f_2^{ijk\hat{i}\hat{j}\hat{k}} : ijk\hat{i}\hat{j}\hat{k} \in \mathcal{S}_2\}$ and $g \in \{f_1^{ijk\hat{i}\hat{j}\hat{k}} : ijk\hat{i}\hat{j}\hat{k} \in \mathcal{S}_1\}$. Let $f = f_2^{ijk\hat{i}j\hat{k}} = -x_{ijk}x_{\hat{i}j\hat{k}} + x_{\hat{i}jk}x_{ij\hat{k}}$ and $g = f_1^{ijk\hat{i}\hat{j}k} = -x_{ijk}x_{\hat{i}\hat{j}k} + x_{\hat{i}jk}x_{i\hat{j}k}$ for some $ijk\hat{i}\hat{j}\hat{k} \in \mathcal{S}_0$. Then

$$S(f,g) = x_{ijk}\left(-x_{\hat{i}\hat{j}k}x_{\hat{i}jk} + x_{ij\hat{k}}x_{\hat{i}\hat{j}k}\right) = x_{ijk}\left(-f_1^{ijk\hat{i}\hat{j}\hat{k}} + f_3^{ijk\hat{i}\hat{j}\hat{k}}\right) \to_{\mathcal{G}_3} 0.$$

For the remaining cases one proceeds similarly. In order to show that $\mathcal{G}_3$ is the reduced Gröbner basis, one uses the same arguments as in the proof of Lemma 2. □

Remark 3. The above Gröbner basis $\mathcal{G}_3$ is obtained by taking all minors of order two of all three unfoldings of the tensor $\mathbf{X} \in \mathbb{R}^{n_1\times n_2\times n_3}$ (not considering the same minor twice). One might think that the θ_1-norm obtained in this way corresponds to a (weighted) sum of the nuclear norms of the unfoldings, which has been used in [20, 31] for tensor recovery. The examples of cubic tensors $\mathbf{X} \in \mathbb{R}^{2\times 2\times 2}$ presented in Table 3 show that this is not the case. Assuming that θ_1-norm is a linear combination of the nuclear norm of the unfoldings, there exist $\alpha, \beta, \gamma \in \mathbb{R}$ such that $\alpha\|\mathbf{X}^{\{1\}}\|_* + \beta\|\mathbf{X}^{\{2\}}\|_* + \gamma\|\mathbf{X}^{\{3\}}\|_* = \|\mathbf{X}\|_{\theta_1}$. From the first and the second tensor in Table 3 we obtain $\gamma = 0$. Similarly, the first and the third tensor, and the first and fourth tensor give $\beta = 0$ and $\alpha = 0$, respectively. Thus, the θ_1-norm does not coincide with a weighted sum of the nuclear norms of the unfoldings. In addition, the last tensor shows that the θ_1-norm does not equal maximum of the norms of the unfoldings.

4.2. **Fourth order tensors.** For the case of fourth order tensors, we have several possibilities of defining polynomial ideals generated by order two minors whose algebraic variety is the set of rank one tensors of Frobenius norm one. Indeed, we can choose the minors corresponding to any set of matricizations such that the tensor is of rank one if and only if all order two minors of the matricization vanish. Each set of matricizations possibly defines a different family of θ_k-norms. Note that different notions of tensor rank and tensor format are associated to different sets of matricizations. We will first discuss the set of all unfoldings and the related θ_1-norm and then the set of all matricizations and the corresponding θ_k-norm. Other tensor formats [28] – e.g. the tensor train decomposition [41] – and their associated θ_k-norms are left to future investigations.

The unfolding theta norm. We start by introducing some notation. We denote by I the subscript $i_1i_2i_3i_4$ and similarly write $\hat{I} = \hat{i}_1\hat{i}_2\hat{i}_3\hat{i}_4$. For $\mathcal{S} \subset [d]$ we introduce

$$\mathcal{D}^{\mathcal{S}} := \{(I,\hat{I}) : i_k < \hat{i}_k, \text{ for all } k \in \mathcal{S},\ i_\ell \leq \hat{i}_\ell, \text{ for all } \ell \notin \mathcal{S}\}$$
$$\mathcal{E}^{\mathcal{S}} := \{(I,\hat{I}) : i_k < \hat{i}_k, \text{ for all } k \in \mathcal{S},\ i_\ell = \hat{i}_\ell, \text{ for all } \ell \notin \mathcal{S}\}.$$

As already noted, a tensor is a rank one tensor if and only if all its unfoldings are rank one matrices. Therefore, the following set of polynomials (corresponding to the order two minors of all unfoldings)

$$f^{\{1\}}_{(I,\hat{I})}(\mathbf{x}) = -x_{i_1i_2i_3i_4}x_{\hat{i}_1\hat{i}_2\hat{i}_3\hat{i}_4} + x_{i_1\hat{i}_2\hat{i}_3\hat{i}_4}x_{\hat{i}_1i_2i_3i_4}, \quad (I,\hat{I}) \in \mathcal{S}_1^4 := \mathcal{D}^{\{1\}}\backslash\mathcal{E}^{\{1\}}$$
$$f^{\{2\}}_{(I,\hat{I})}(\mathbf{x}) = -x_{i_1i_2i_3i_4}x_{\hat{i}_1\hat{i}_2\hat{i}_3\hat{i}_4} + x_{i_1\hat{i}_2i_3i_4}x_{\hat{i}_1i_2\hat{i}_3\hat{i}_4}, \quad (I,\hat{I}) \in \mathcal{S}_2^4 = \mathcal{D}^{\{2\}}\backslash\{\mathcal{E}^{\{1\}} \cup \mathcal{E}^{\{2\}}\}$$
$$f^{\{3\}}_{(I,\hat{I})}(\mathbf{x}) = -x_{i_1i_2i_3i_4}x_{\hat{i}_1\hat{i}_2\hat{i}_3\hat{i}_4} + x_{i_1i_2\hat{i}_3i_4}x_{\hat{i}_1\hat{i}_2i_3\hat{i}_4}, \quad (I,\hat{I}) \in \mathcal{S}_3^4 = \mathcal{D}^{\{3\}}\backslash\{\mathcal{E}^{\{1\}} \cup \mathcal{E}^{\{2\}} \cup \mathcal{E}^{\{3\}}\}$$
$$f^{\{4\}}_{(I,\hat{I})}(\mathbf{x}) = -x_{i_1i_2i_3i_4}x_{\hat{i}_1\hat{i}_2\hat{i}_3\hat{i}_4} + x_{i_1i_2i_3\hat{i}_4}x_{\hat{i}_1\hat{i}_2\hat{i}_3i_4}, \quad (I,\hat{I}) \in \mathcal{S}_4^4 = \mathcal{D}^{\{4\}}\backslash\{\mathcal{E}^{\{1\}} \cup \mathcal{E}^{\{2\}} \cup \mathcal{E}^{\{3\}} \cup \mathcal{E}^{\{4\}}\}$$
$$g_4(\mathbf{x}) = \sum_{i_1=1}^{n_1}\sum_{i_2=1}^{n_2}\sum_{i_3=1}^{n_3}\sum_{i_4=1}^{n_4} x^2_{i_1i_2i_3i_4} - 1, \tag{26}$$

generates a polynomial ideal $J_{u,4}$ in $\mathbb{R}[\mathbf{x}] = \mathbb{R}[x_{1111}, x_{1112}, \ldots, x_{n_1n_2n_3n_4}]$ such that its algebraic variety $\nu_{\mathbb{R}}(J_{u,4})$ coincides with the set of rank one Frobenius norm one fourth order tensors. Unfortunately, the set of generators

$$\mathcal{H}_{u,4} = \left\{f^{\{1\}}_{(I,\hat{I})} : (I,\hat{I}) \in \mathcal{S}_1^4\right\} \cup \left\{f^{\{2\}}_{(I,\hat{I})} : (I,\hat{I}) \in \mathcal{S}_2^4\right\} \cup \left\{f^{\{3\}}_{(I,\hat{I})} : (I,\hat{I}) \in \mathcal{S}_3^4\right\} \cup \left\{f^{\{4\}}_{(I,\hat{I})} : (I,\hat{I}) \in \mathcal{S}_4^4\right\} \cup \{g_4\}$$

does not form a Gröbner basis of the ideal $J_{u,4}$ with respect to the grevlex ordering. To see this, let $f_1, f_2, f_3 \in \mathcal{H}_{u,4}$ be defined as

$$f_1 = -x_{2111}x_{2222} + x_{2211}x_{2122},$$
$$f_2 = -x_{1122}x_{2222} + x_{2122}x_{1222},$$
$$f_3 = -x_{1111}x_{2222} + x_{2111}x_{1222}.$$

Notice that $S(f_1, f_2) = -x_{1222}x_{2111}x_{2222} + x_{1122}x_{2222}x_{2211}$ and after the division by f_3, we obtain $S(f_1, f_2) = -x_{2222}f_3 + r$, where

$$r = -x_{1111}x^2_{2222} + x_{1122}x_{2211}x_{2222}.$$

However, no monomial of r is divisible by any of the leading monomials in $\mathcal{H}_{u,4}$. In other words, we found two polynomials $f_1, f_2 \in \mathcal{H}_{u,4}$ such that $S(f_1, f_2) \not\to_{\mathcal{H}_{u,4}} 0$, i.e., the set $\mathcal{H}_{u,4}$ does not form a Gröbner basis with respect to the grevlex ordering.

Applying Buchberger's algorithm we can extend the generating set $\mathcal{H}_{u,4}$ to the reduced Gröbner basis $\mathcal{G}_{u,4}$ of the ideal $J_{u,4}$ (for details about this procedure see [13, 12]). Notice that

$$\deg(f) \geq 3, \qquad \text{for all } f \in \mathcal{G}_{u,4}\backslash\mathcal{H}_{u,4}$$

since there are no two polynomials in $\mathcal{H}_{u,4}$ that have the same leading term.

However, in order to compute the first theta body $\mathrm{TH}_1(J_{u,4})$, we do not need to compute the full reduced Gröbner basis $\mathcal{G}_{u,4}$ because this requires only the θ-basis reduced to degree two. The first theta body $\mathrm{TH}_1(J_{u,4})$ defines a new norm which we call the *unfolding*-θ_1-norm denoted as $\|\cdot\|_{u,\theta_1}$. Again, this norm cannot be written as the convex combination of the nuclear norms of the unfoldings (see Table 4).

The full theta norm. Next, we consider all matricizations in order to define another theta norm for fourth order tensors, which we call the *full* θ_k-norm. Again, we use the fact a tensor is rank one if and only if all its matricizations (not just the unfoldings) are rank one matrices. This leads to the following polynomials (completing the set $\mathcal{H}_{u,4}$), each one corresponding to a specific matricization,

$$\begin{aligned}
\mathbf{X}^{\{1\}}: f^{\{1\}}_{(I,\hat{I})}(\mathbf{x}) &= -x_{i_1 i_2 i_3 i_4} x_{\hat{i}_1 \hat{i}_2 \hat{i}_3 \hat{i}_4} + x_{i_1 \hat{i}_2 \hat{i}_3 \hat{i}_4} x_{\hat{i}_1 i_2 i_3 i_4}, \quad (I,\hat{I}) \in \mathcal{S}_1^4 \\
\mathbf{X}^{\{2\}}: f^{\{2\}}_{(I,\hat{I})}(\mathbf{x}) &= -x_{i_1 i_2 i_3 i_4} x_{\hat{i}_1 \hat{i}_2 \hat{i}_3 \hat{i}_4} + x_{\hat{i} x_1 i_2 \hat{i}_3 \hat{i}_4} x_{i_1 \hat{i}_2 i_3 i_4}, \quad (I,\hat{I}) \in \mathcal{S}_2^4 \\
\mathbf{X}^{\{3\}}: f^{\{3\}}_{(I,\hat{I})}(\mathbf{x}) &= -x_{i_1 i_2 i_3 i_4} x_{\hat{i}_1 \hat{i}_2 \hat{i}_3 \hat{i}_4} + x_{\hat{i}_1 \hat{i}_2 i_3 \hat{i}_4} x_{i_1 i_2 \hat{i}_3 i_4}, \quad (I,\hat{I}) \in \mathcal{S}_3^4 \\
\mathbf{X}^{\{4\}}: f^{\{4\}}_{(I,\hat{I})}(\mathbf{x}) &= -x_{i_1 i_2 i_3 i_4} x_{\hat{i}_1 \hat{i}_2 \hat{i}_3 \hat{i}_4} + x_{\hat{i}_1 \hat{i}_2 \hat{i}_3 i_4} x_{i_1 i_2 i_3 \hat{i}_4}, \quad (I,\hat{I}) \in \mathcal{S}_4^4 \\
\mathbf{X}^{\{1,2\}}: f^{\{1,2\}}_{(I,\hat{I})}(\mathbf{x}) &= -x_{i_1 i_2 i_3 i_4} x_{\hat{i}_1 \hat{i}_2 \hat{i}_3 \hat{i}_4} + x_{i_1 i_2 \hat{i}_3 \hat{i}_4} x_{\hat{i}_1 \hat{i}_2 i_3 i_4}, \quad (I,\hat{I}) \in \mathcal{S}_0 \\
\mathbf{X}^{\{1,3\}}: f^{\{1,3\}}_{(I,\hat{I})}(\mathbf{x}) &= -x_{i_1 i_2 i_3 i_4} x_{\hat{i}_1 \hat{i}_2 \hat{i}_3 \hat{i}_4} + x_{i_1 \hat{i}_2 i_3 \hat{i}_4} x_{\hat{i}_1 i_2 \hat{i}_3 i_4}, \quad (I,\hat{I}) \in \mathcal{S}_0 \\
\mathbf{X}^{\{1,4\}}: f^{\{1,4\}}_{(I,\hat{I})}(\mathbf{x}) &= -x_{i_1 i_2 i_3 i_4} x_{\hat{i}_1 \hat{i}_2 \hat{i}_3 \hat{i}_4} + x_{i_1 \hat{i}_2 \hat{i}_3 i_4} x_{\hat{i}_1 i_2 i_3 \hat{i}_4}, \quad (I,\hat{I}) \in \mathcal{S}_0 \\
g_4(\mathbf{x}) &= \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \sum_{i_3=1}^{n_3} \sum_{i_4=1}^{n_4} x^2_{i_1 i_2 i_3 i_4} - 1
\end{aligned} \tag{27}$$

where $\mathcal{S}_0 = \left\{ i < \hat{i}, j < \hat{j}, k < \hat{k}, l < \hat{l} \right\}$ and the sets $\mathcal{S}_1^4$, $\mathcal{S}_2^4$, $\mathcal{S}_3^4$, and $\mathcal{S}_4^4$ are defined as in (26). We then define J_4 as the polynomial ideal in $\mathbb{R}[\mathbf{x}] = \mathbb{R}[x_{1111}, x_{1112}, \ldots, x_{n_1 n_2 n_3 n_4}]$ generated by these polynomials, i.e.,

$$\begin{aligned}
J_4 = \langle \mathcal{G}_4 \rangle = \Big\langle &\{f^{\{1\}}_{(I,\hat{I})}\}_{(I,\hat{I}) \in \mathcal{S}_1^4} \cup \{f^{\{2\}}_{(I,\hat{I})}\}_{(I,\hat{I}) \in \mathcal{S}_2^4} \cup \{f^{\{3\}}_{(I,\hat{I})}\}_{(I,\hat{I}) \in \mathcal{S}_3^4} \cup \{f^{\{4\}}_{(I,\hat{I})}\}_{(I,\hat{I}) \in \mathcal{S}_4^4} \\
&\cup \{f^{\{1,2\}}_{(I,\hat{I})}\}_{(I,\hat{I}) \in \mathcal{S}_0} \cup \{f^{\{1,3\}}_{(I,\hat{I})}\}_{(I,\hat{I}) \in \mathcal{S}_0} \cup \{f^{\{1,4\}}_{(I,\hat{I})}\}_{(I,\hat{I}) \in \mathcal{S}_0} \cup g_4 \Big\rangle,
\end{aligned}$$

Note that the above set of polynomials does not include all second order minors for all matricizations. For instance, the polynomial

$$h(\mathbf{x}) = -x_{1234} x_{2343} + x_{1243} x_{2334}$$

which corresponds to a minor of the matricization $\mathbf{X}^{\{1,2\}}$ does not belong to the basis $\mathcal{G}_4$. However, $h(\mathbf{x})$ is in the ideal J_4 since $h = f - g$ with

$$\begin{aligned}
f(\mathbf{x}) &= -x_{1233} x_{2344} + x_{1243} x_{2334} \in \mathcal{G}_4 \\
\text{and } g(\mathbf{x}) &= -x_{1233} x_{2344} + x_{1234} x_{2343} \in \mathcal{G}_4.
\end{aligned}$$

In fact, all possible minors of all possible matricizations belong to the ideal J_4 (and can be expressed similarly as above as a difference of two polynomials from the basis $\mathcal{G}_4$), but to define the reduced Gröbner basis, it is enough to consider the generating set $\mathcal{G}_4$.

Theorem 4. *The set $\mathcal{G}_4$ forms the reduced Gröbner basis of the ideal J_4.*

Proof. The statement is a special case of Theorem 5 below concerning general d-th order tensors. □

	$\mathbf{X} \in \mathbb{R}^{2\times 2\times 2\times 2}$	$\Vert\mathbf{X}^{\{1\}}\Vert_*$	$\Vert\mathbf{X}^{\{2\}}\Vert_*$	$\Vert\mathbf{X}^{\{3\}}\Vert_*$	$\Vert\mathbf{X}^{\{4\}}\Vert_*$	$\Vert\mathbf{X}\Vert_{\theta_1}$	$\Vert\mathbf{X}\Vert_{u,\theta_1}$
1	$\mathbf{X}(:,:,:,1) = \left[\begin{array}{cc\|cc} 1 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 \end{array}\right]$ $\mathbf{X}(:,:,:,2) = \left[\begin{array}{cc\|cc} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{array}\right]$	$1+\sqrt{6}$	$2+\sqrt{3}$	$\sqrt{2}+\sqrt{5}$	$\sqrt{2}+\sqrt{5}$	5	4.2361
2	$\mathbf{X}(:,:,:,1) = \left[\begin{array}{cc\|cc} 1 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 \end{array}\right]$ $\mathbf{X}(:,:,:,2) = \left[\begin{array}{cc\|cc} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{array}\right]$	$1+\sqrt{6}$	$\sqrt{2}+\sqrt{5}$	$2+\sqrt{3}$	$\sqrt{2}+\sqrt{5}$	5	4.2361
3	$\mathbf{X}(:,:,:,1) = \left[\begin{array}{cc\|cc} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right]$ $\mathbf{X}(:,:,:,2) = \left[\begin{array}{cc\|cc} 0 & 1 & 0 & 0 \\ 2 & 0 & 0 & 0 \end{array}\right]$	$2+\sqrt{3}$	$\sqrt{2}+\sqrt{5}$	$1+\sqrt{6}$	$\sqrt{2}+\sqrt{5}$	5	4.2361
4	$\mathbf{X}(:,:,:,1) = \left[\begin{array}{cc\|cc} 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{array}\right]$ $\mathbf{X}(:,:,:,2) = \left[\begin{array}{cc\|cc} 0 & 1 & 0 & 0 \\ 2 & 0 & 0 & 0 \end{array}\right]$	$\sqrt{3}+\sqrt{5}$	$\sqrt{2}+\sqrt{6}$	$\sqrt{2}+\sqrt{6}$	$\sqrt{3}+\sqrt{5}$	6	4.6503
5	$\mathbf{X}(:,:,:,1) = \left[\begin{array}{cc\|cc} 1 & 0 & 0 & 2 \\ 0 & 0 & 1 & 0 \end{array}\right]$ $\mathbf{X}(:,:,:,2) = \left[\begin{array}{cc\|cc} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{array}\right]$	$\sqrt{2}+\sqrt{6}$	$\sqrt{3}+\sqrt{5}$	$\sqrt{3}+\sqrt{5}$	$\sqrt{2}+\sqrt{6}$	6	4.6503
6	$\mathbf{X}(:,:,:,1) = \left[\begin{array}{cc\|cc} 2 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right]$ $\mathbf{X}(:,:,:,2) = \left[\begin{array}{cc\|cc} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{array}\right]$	$1+\sqrt{6}$	$\sqrt{2}+\sqrt{5}$	$1+\sqrt{6}$	$\sqrt{2}+\sqrt{5}$	5	4.4142

TABLE 4. Nuclear norms of unfoldings of fourth order tensors together with their θ_1-norm and unfolding-θ_1-norm which were computed numerically.

4.3. **The theta norm for general dth order tensors.** Let us now consider dth order tensors in $\mathbb{R}^{n_1\times n_2\times\cdots\times n_d}$ for general $d \geq 4$. Our approach relies again on the fact that a tensor $\mathbf{X} \in \mathbb{R}^{n_1\times n_2\times\cdots\times n_d}$ is of rank one if and only if all its matricizations are rank one matrices, or equivalently, if all minors of order two of each matricization vanish.

The description of the polynomial ideal generated by the second order minors of all matricizations of a tensor $\mathbf{X} \in \mathbb{R}^{n_1\times n_2\times\cdots\times n_d}$ unfortunately requires some technical notation. Again, we do not need all such minors in the generating set that we introduce next. In fact, this generating set will turn out to be the reduced Gröbner basis of the ideal.

Similarly as before, the entry $(i_1, i_2, \ldots, i_d)$ of a tensor $\mathbf{X} \in \mathbb{R}^{n_1\times n_2\times\cdots\times n_d}$ corresponds to the variable $x_{i_1 i_2\cdots i_d}$. We aim at introducing a set of polynomials of the form

$$f_{d,(I,\hat{I})}^{\mathcal{S},\mathcal{M}} := -x_I x_{\hat{I}_{\mathcal{S}}} + x_{I_{\mathcal{S},\mathcal{M}}} x_{\hat{I}_{\mathcal{S},\mathcal{M}}} \tag{28}$$

which will generate the desired polynomial ideal. These polynomials correspond to a minor of a matricization $\mathbf{X}^{\mathcal{M}}$ – thus, $\{I(k), \hat{I}_{\mathcal{S}}(k)\} = \{I_{\mathcal{S},\mathcal{M}}(k), \hat{I}_{\mathcal{S},\mathcal{M}}(k)\}$, for every $k \in [d]$, where $I(k)$ denotes the k-th entry of I. For instance, if $I = 1\,3\,5\,7\,9$, then $I(3) = 5$. The set $\mathcal{S}$ denotes the indices where I and $\hat{I}_{\mathcal{S}}$ differ. Since for a minor of order two of a matricization $\mathbf{X}^{\mathcal{M}}$ the sets I and $\hat{I}_{\mathcal{S}}$ need to differ in at least two indices, $\mathcal{S}$ is contained in

$$\mathcal{S}_{[d]} := \{\mathcal{S} \subset [d] : 2 \leq |\mathcal{S}| \leq d\}.$$

For two multiindices $I, \hat{I}$ and $\mathcal{S} \in \mathcal{S}_{[d]}$, we define a monomial $X_{\hat{I}_{\mathcal{S}}}$ with subscripts

$$\hat{I}_{\mathcal{S}}(k) = \begin{cases} i_k, & \text{if } k \notin \mathcal{S}, \\ \hat{i}_k, & \text{if } k \in \mathcal{S}. \end{cases}$$

Given the set $\mathcal{S}$ of differing indices, we require all subsets $\mathcal{M} \subseteq \mathcal{S}$ of possible indices which are "switched" between and I and $I_{\mathcal{S},\mathcal{M}}$ for forming the minors in (28). The set $\mathcal{M}$ corresponds to an associated matricization $\mathbf{X}^{\mathcal{M}}$. The set of possible subsets $\mathcal{M}$ is given as

$$\mathcal{P}_{\mathcal{S}} = \begin{cases} \left\{\mathcal{M} \subset \mathcal{S} : |\mathcal{M}| \leq \lfloor \frac{|\mathcal{S}|}{2} \rfloor \right\} \backslash \{\emptyset\}, & \text{if } |\mathcal{S}| \text{ is odd,} \\ \left\{\mathcal{M} \subset \mathcal{S} : |\mathcal{M}| \leq \lfloor \frac{|\mathcal{S}|-1}{2} \rfloor \right\} \cup \left\{\mathcal{M} \subset \mathcal{S} : |\mathcal{M}| = \frac{|\mathcal{S}|}{2}, 1 \in \mathcal{M}\right\} \backslash \{\emptyset\}, & \text{if } |\mathcal{S}| \text{ is even.} \end{cases}$$

Notice that $\mathcal{P}_{\mathcal{S}} \cup \mathcal{P}_{\mathcal{S}^c} \cup \{\emptyset\} \cup \mathcal{S}$ with $\mathcal{P}_{\mathcal{S}^c} := \{\mathcal{M} : \mathcal{S} \backslash \mathcal{M} \in \mathcal{P}_{\mathcal{S}}\}$ forms the power set of $\mathcal{S}$. The constraint on the size of $\mathcal{M}$ in the definition of $\mathcal{P}_{\mathcal{S}}$ is motivated by the fact that the role of I and $I_{\mathcal{S},\mathcal{M}}$ can be switched and only lead to the negative of the polynomial $f_{d,(I,\hat{I})}^{\mathcal{S},\mathcal{M}}$ below.

Next, we define the monomials $X_{I_{\mathcal{S},\mathcal{M}}} X_{\hat{I}_{\mathcal{S},\mathcal{M}}}$, $\mathcal{M} \in \mathcal{P}_{\mathcal{S}}$, with the corresponding subscripts

$$I_{\mathcal{S},\mathcal{M}}(k) = \begin{cases} \hat{i}_k, & \text{if } k \in \mathcal{M} \\ i_k, & \text{if } k \in \mathcal{S} \backslash \mathcal{M} \end{cases} \quad \text{and} \quad \hat{I}_{\mathcal{S},\mathcal{M}}(k) = \begin{cases} i_k, & \text{if } k \in \mathcal{M} \\ \hat{i}_k, & \text{if } k \in \mathcal{S} \backslash \mathcal{M} \end{cases}.$$

Finally, for fixed $\mathcal{S} \in \mathcal{S}_{[d]}$ we introduce the polynomials

$$f_{d,(I,\hat{I})}^{\mathcal{S},\mathcal{M}}(\mathbf{x}) := -x_I x_{\hat{I}_{\mathcal{S}}} + x_{I_{\mathcal{S},\mathcal{M}}} x_{\hat{I}_{\mathcal{S},\mathcal{M}}} \quad \text{for } \mathcal{M} \in \mathcal{P}_{\mathcal{S}} \text{ and } (I, \hat{I}) \in \mathcal{T}_d^{\mathcal{S}},$$

where

$$\mathcal{T}_d^{\mathcal{S}} = \{(I, \hat{I}) : i_k < \hat{i}_k, \text{ for all } k \in \mathcal{S} \text{ and } i_\ell = \hat{i}_\ell, \text{ for all } \ell \notin \mathcal{S}\}. \tag{29}$$

For notational purposes, we define

$$\{f_d^{\mathcal{S}}\} = \{f_{d,(I,\hat{I})}^{\mathcal{S},\mathcal{M}} : \mathcal{M} \in \mathcal{P}_{\mathcal{S}}, (I, \hat{I}) \in \mathcal{T}_d^{\mathcal{S}}\} \quad \text{for } \mathcal{S} \in \mathcal{S}_{[d]}.$$

Since we are interested in Frobenius norm one tensors, we also introduce the polynomial

$$g_d(\mathbf{x}) = \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \ldots \sum_{i_d=1}^{n_d} x_{i_1 i_2 \ldots i_d}^2 - 1.$$

Our polynomial ideal is then the one generated by the polynomials in

$$\mathcal{G}_d = \bigcup_{\mathcal{S} \in \mathcal{S}_{[d]}} \{f_d^{\mathcal{S}}\} \cup \{g_d\} \subset \mathbb{R}[\mathbf{x}] = \mathbb{R}[x_{11\ldots1}, x_{11\ldots2}, \ldots, x_{n_1 n_2 \ldots n_d}], \quad \text{i.e.,} \quad J_d = \langle \mathcal{G}_d \rangle.$$

As in the special case of the fourth order case, not all second order minors corresponding to all matricizations are contained in the generating set $\mathcal{G}_d$ due to the condition $i_k < \hat{i}_k$ for all $k \in \mathcal{S}$ in the definition of $\mathcal{T}_d^{\mathcal{S}}$. Nevertheless all second order minors are contained in the ideal J_d as will also be revealed by the proof of Theorem 5 below. For instance, $h(\mathbf{x}) = -x_{1234}x_{2343} + x_{1243}x_{2334}$ – corresponding to a minor of the matricization $\mathbf{X}^{\mathcal{M}}$ for $\mathcal{M} = \{1, 2\}$ – does not belong to $\mathcal{G}_4$, but it does belong to the ideal J_4. Moreover, it is straightforward to verify that all polynomials in $\mathcal{G}_d$ differ from each other.

The algebraic variety of J_d consists of all rank one Frobenius norm one order d tensors as desired, and its convex hull yields the tensor nuclear norm ball.

Theorem 5. *The set $\mathcal{G}_d$ forms the reduced Gröbner basis of the ideal J_d with respect to the grevlex order.*

Proof of Theorem 5. Again, we will use Buchberger's criterion stated in Theorem 6. First notice that the polynomials g_d and $f_{d,(I,\hat{I})}^{\mathcal{S},\mathcal{M}}$ are always relatively prime, since $\mathrm{LM}(g_d) = x_{11\cdots1}^2$ and $\mathrm{LM}(f_{d,(I,\hat{I})}^{\mathcal{S},\mathcal{M}}) = x_{I_{\mathcal{S},\mathcal{M}}} x_{\hat{I}_{\mathcal{S},\mathcal{M}}}$ for $\mathcal{S} \in \mathcal{S}_{[d]}$, $\mathcal{M} \in \mathcal{P}_{\mathcal{S}}$ and $(I, \hat{I}) \in \mathcal{T}_d^{\mathcal{S}}$. Therefore, we need to show that $S(f_1, f_2) \to_{\mathcal{G}_d} 0$, for all $f_1, f_2 \in \mathcal{G}_d \backslash \{g_d\}$ with $f_1 \neq f_2$. To this end, we analyze the division algorithm on $\langle \mathcal{G}_d \rangle$.

Let $f_1, f_2 \in \mathcal{G}_d$ with $f_1 \neq f_2$. Then it holds $\mathrm{LM}(f_1) \neq \mathrm{LM}(f_2)$. If these leading terms are not relatively prime, the S-polynomial is of the form

$$S(f_1, f_2) = x_{I_1} x_{I_2} x_{I_3} - x_{\bar{I}_1} x_{\bar{I}_2} x_{\bar{I}_3}$$

with $\{I_1(k), I_2(k), I_3(k)\} = \left\{\bar{I}_1(k), \bar{I}_2(k), \bar{I}_3(k)\right\}$ for all $k \in [d]$.

The step-by-step procedure of the division algorithm for our scenario is presented in Table 5. We will show that the algorithm eventually stops and that step 2) is feasible, i.e., that there always exist k and ℓ such that (30) holds – provided that $S^i \neq 0$. (In fact, in some sense the purpose of the algorithm is to achieve the condition $i_k \leq \hat{i}_k$ for all k that appears in the definition (29) of the set of allowed indices in a finite number of steps.) This will show then that $S(f_1, f_2) \to_{\mathcal{G}_d} 0$.

Division algorithm on $\langle \mathcal{G}_d \rangle$

Input: polynomials $f_1, f_2 \in \mathcal{G}_d$

$S^0 = S(f_1, f_2) = x_{I_1} x_{I_2} x_{I_3} - x_{\bar{I}_1} x_{\bar{I}_2} x_{\bar{I}_3}$, $i = 0$

while $S^i \neq 0$ **do**

1) **Let** $\mathrm{LM}(S^i) = x_{\hat{I}_{1,i}} x_{\hat{I}_{2,i}} x_{\hat{I}_{3,i}}$ **and** $\mathrm{NLM}(S^i) = \left| S^i - \mathrm{LT}(S^i) \right|$

2) **Find indices** $I_{1,i}, I_{2,i} \in \{\hat{I}_{1,i}, \hat{I}_{2,i}, \hat{I}_{3,i}\}$ **such that there exist at least one** k **and at least one** ℓ **for which**

$$I_{1,i}(k) < I_{2,i}(k) \quad \text{and} \quad I_{1,i}(\ell) > I_{2,i}(\ell) \tag{30}$$

and let $I_{3,i}$ **be the remaining index in** $\{\hat{I}_{1,i}, \hat{I}_{2,i}, \hat{I}_{3,i}\} \backslash \{I_{1,i}, I_{2,i}\}$.

3) **Define**

$$\mathcal{S}_{1,i} := \{k \in [d] : I_{1,i}(k) < I_{2,i}(k)\}, \quad \mathcal{S}_{2,i} := \{\ell \in [d] : I_{1,i}(\ell) > I_{2,i}(\ell)\},$$
$$\mathcal{S}_i := \mathcal{S}_{1,i} \cup \mathcal{S}_{2,i}.$$

If $|\mathcal{S}_i|$ **is odd set** $\mathcal{M}_i := \begin{cases} \mathcal{S}_{1,i}, & \textbf{if } |\mathcal{S}_{1,i}| \leq \lfloor \frac{|\mathcal{S}_i|}{2} \rfloor, \\ \mathcal{S}_{2,i}, & \textbf{if } |\mathcal{S}_{2,i}| \leq \lfloor \frac{|\mathcal{S}_i|}{2} \rfloor. \end{cases}$

If $|\mathcal{S}_i|$ **is even set** $\mathcal{M}_i := \begin{cases} \mathcal{S}_{1,i}, & \textbf{if } |\mathcal{S}_{1,i}| \leq \lfloor \frac{|\mathcal{S}_i|-1}{2} \rfloor \textbf{ or } |\mathcal{S}_{1,i}| = \frac{|\mathcal{S}_i|}{2} \textbf{ and } 1 \in \mathcal{S}_{1,i}, \\ \mathcal{S}_{2,i}, & \textbf{if } |\mathcal{S}_{2,i}| \leq \lfloor \frac{|\mathcal{S}_i|-1}{2} \rfloor \textbf{ or } |\mathcal{S}_{2,i}| = \frac{|\mathcal{S}_i|}{2} \textbf{ and } 1 \in \mathcal{S}_{2,i}. \end{cases}$

Define

$$I_{\mathcal{M}_i}(k) := \begin{cases} I_{1,i}(k) & \textbf{if } k \in \mathcal{S}_{1,i} \\ I_{2,i}(k) & \textbf{if } k \notin \mathcal{S}_{1,i} \end{cases} \quad \textbf{and} \quad I_{\mathcal{M}_i^c}(k) := \begin{cases} I_{1,i}(k) & \text{if } k \notin \mathcal{S}_{1,i} \\ I_{2,i}(k) & \textbf{if } k \in \mathcal{S}_{1,i} \end{cases}.$$

4) **Divide** S^i **by** $f^{\mathcal{S}_i, \mathcal{M}_i}_{d,\left(I_{\mathcal{M}_i}, I_{\mathcal{M}_i^c}\right)} = x_{I_{1,i}} x_{I_{2,i}} - x_{I_{\mathcal{M}_i}} x_{I_{\mathcal{M}_i^c}}$ **to obtain**

$$S^i = \mathrm{LC}(S^i) \left[x_{I_{3,i}} (-x_{I_{\mathcal{M}_i}} x_{I_{\mathcal{M}_i^c}} + x_{I_{1,i}} x_{I_{2,i}}) + x_{I_{\mathcal{M}_i}} x_{I_{\mathcal{M}_i^c}} x_{I_{3,i}} - \mathrm{NLM}(S^i) \right].$$

5) **Define**

$$S^{i+1} := x_{I_{\mathcal{M}_i}} x_{I_{\mathcal{M}_i^c}} x_{I_{3,i}} - \mathrm{NLM}(S^i).$$

6) $i = i + 1$

end while

TABLE 5. The division algorithm on the ideal $\langle \mathcal{G}_d \rangle$.

Before passing to the general proof, we illustrate the division algorithm on an example for $d = 4$. The experienced reader may skip this example.

Let $f_1 = f^{[4],\{2\}}_{4,(1112,2223)} = -x_{1112} x_{2223} + x_{1212} x_{2123} \in \mathcal{G}_4$ and $f_2 = f^{[4],\{1,2\}}_{4,(2111,3323)} = -x_{2111} x_{3323} + x_{2123} x_{3311} \in \mathcal{G}_4$. We will show that $S(f_1, f_2) = -x_{1112} x_{2223} x_{3311} + x_{1212} x_{2111} x_{3323} \to_{\mathcal{G}_4} 0$ by going through the division algorithm.

In iteration $i = 0$ we set $S^0 = S(f_1, f_2) = -x_{1112} x_{2223} x_{3311} + x_{1212} x_{2111} x_{3323}$. The leading monomial is $\mathrm{LM}(S^0) = x_{1112} x_{2223} x_{3311}$, the leading coefficient $\mathrm{LC}(S^0) = -1$ and the non-leading monomial $\mathrm{NLM}(S^0) = x_{1212} x_{2111} x_{3323}$. Among the two options for choosing a pair of indexes $(I_{1,0}, I_{2,0})$ in step 2), we decide to take $I_{1,0} = 1112$ and $I_{2,0} = 3311$ such that $\mathcal{S}^1 = \{1, 2\}$ and $\mathcal{M}_0 = \mathcal{S}^2 = \{4\}$. The polynomial $x_{I_{1,0}} x_{I_{2,0}} - x_{I_{\mathcal{M}_0}} x_{I_{\mathcal{M}_0^c}}$ then equals the polynomial

$f_{4,(1111,3312)}^{\{1,2,4\},\{4\}} = -x_{1111}x_{3312} + x_{1112}x_{3311} \in \mathcal{G}_4$ and we can write

$$S^0 = -1 \cdot \Big(x_{2223}\left(-x_{1111}x_{3312} + x_{1112}x_{3311}\right) + \underbrace{x_{1111}x_{2223}x_{3312} - x_{1212}x_{2111}x_{3323}}_{=S^1} \Big).$$

The leading and non-leading monomials of S^1 are $\mathrm{LM}(S^1) = x_{1111}x_{2223}x_{3312}$ and $\mathrm{NLM}(S^1) = x_{1212}x_{2111}x_{3323}$, respectively, while $\mathrm{LC}(S^1) = 1$. The only option for a pair of indices as in (30) is $I_{1,1} = 2223, I_{2,1} = 3312$, so that $\mathcal{S}^1 = \{1,2\}$, $\mathcal{S}^2 = \{3,4\}$ and $\mathcal{M}_1 = \mathcal{S}^1$. The divisor $x_{I_{1,1}}x_{I_{2,1}} - x_{I_{\mathcal{M}_1}}x_{I_{\mathcal{M}_1^c}}$ in the step 4) equals $f_{4,(2212,3323)}^{[4],\{1,2\}} = -x_{2212}x_{3323} + x_{2223}x_{3312} \in \mathcal{G}_4$ and we obtain

$$S^1 = 1 \cdot \Big(x_{1111}\left(-x_{2212}x_{3323} + x_{2223}x_{3312}\right) + \underbrace{x_{1111}x_{2212}x_{3323} - x_{1212}x_{2111}x_{3323}}_{=S_2} \Big).$$

The index sets of the monomial $x_{I_1}x_{I_2}x_{I_3} = x_{1111}x_{2212}x_{3323}$ in S^2 satisfy

$$I_1(k) \le I_2(k) \le I_3(k), \qquad \text{for all } k \in [4]$$

and therefore it is the non-leading monomial of S^2, i.e. $\mathrm{NLM}(S^2) = x_{1111}x_{2212}x_{3323}$. Thus, $\mathrm{LM}(S^2) = x_{1212}x_{2111}x_{3323}$ and $\mathrm{LC}(S^2(f_1,f_2)) = -1$. Now the only option for a pair of indices as in step (30) is $I_{1,2} = 1212$, $I_{2,2} = 2111$ with $\mathcal{S}^1 = \{1\}$, $\mathcal{S}^2 = \{2,4\}$ and $\mathcal{M}_2 = \mathcal{S}^1$. This yields

$$S^2 = -1 \cdot \Big(x_{3323}\left(-x_{1111}x_{2212} + x_{1212}x_{2111}\right) + \underbrace{x_{1111}x_{2212}x_{3323} - x_{1111}x_{2212}x_{3323}}_{=S^3=0} \Big).$$

Thus the division algorithm stops and we obtained after three steps

$$\begin{aligned} S(f_1,f_2) = S^0 &= \mathrm{LC}(S^0)x_{2223}f_{4,(1111,3312)}^{\{1,2,4\},\{4\}} + \mathrm{LC}(S^0)\,\mathrm{LC}(S^1)x_{1111}f_{4,(2212,3323)}^{[4],\{1,2\}} \\ &+ \mathrm{LC}(S^0)\,\mathrm{LC}(S^1)\,\mathrm{LC}(S^2)x_{3323}f_{4,(1111,2212)}^{\{1,2,4\},\{1\}}. \end{aligned}$$

Thus, $S(f_1,f_2) \to_{\mathcal{G}_4} 0$.

Let us now return to the general proof and first show that there always exist indices $I_{1,i}, I_{2,i}$ satisfying (30) unless $S^i = 0$. We start by setting $\mathbf{x}^{\alpha_i} = x_{\hat{I}_{1,i}}x_{\hat{I}_{2,i}}x_{\hat{I}_{3,i}}$ with $x_{\hat{I}_{1,i}} \ge x_{\hat{I}_{2,i}} \ge x_{\hat{I}_{3,i}}$ to be the leading monomial and $\mathbf{x}^{\beta_i}$ be the non-leading monomial of S^i. The existence of a polynomial $h \in \mathcal{G}_d$ such that $\mathrm{LM}(h)$ divides $\mathrm{LM}(S^i) = x_{\hat{I}_{1,i}}x_{\hat{I}_{2,i}}x_{\hat{I}_{3,i}} = \mathbf{x}^{\alpha_{\hat{I}_i}}$ is equivalent to the existence of $I_{1,i}, I_{2,i} \in \left\{\hat{I}_1, \hat{I}_2, \hat{I}_3\right\}$ such that there exists at least one k and at least one ℓ for which $I_{1,i}(k) < I_{2,i}(k)$ and $I_{1,i}(\ell) > I_{2,i}(\ell)$. If such pair does not exist in iteration i, we have

$$\hat{I}_{1,i}(k) \le \hat{I}_{2,i}(k) \le \hat{I}_{3,i}(k) \qquad \text{for all } k \in [d]\,. \tag{31}$$

We claim that this cannot happen if $S^i \neq 0$. In fact, (31) would imply that the monomial $\mathbf{x}^{\alpha_i} = x_{\hat{I}_{1,i}}x_{\hat{I}_{2,i}}x_{\hat{I}_{3,i}}$ is the smallest monomial $x_I x_J x_L$ (with respect to the grevlex order) which satisfies

$$\{I(k), J(k), L(k)\} = \{\hat{I}_{1,i}(k), \hat{I}_{2,i}(k), \hat{I}_{3,i}(k)\} \quad \text{for all } k \in [d]\,.$$

However, then $\mathbf{x}^{\alpha_i}$ would not be the leading monomial by definition of the grevlex order, which leads to a contradiction. Hence, we can always find indices $I_{1,i}, I_{2,i}$ satisfying (30) in step 2) unless $S^i = 0$.

Next we show that the division algorithm always stops in a finite number of steps. We start with iteration $i = 0$ and assume that $S^0 \neq 0$. We choose $I_{1,0}, I_{2,0}, I_{3,0}$ as in (30). Then we divide the polynomial $\mathrm{LM}(S^0) = x_{\hat{I}_{1,0}}x_{\hat{I}_{2,0}}x_{\hat{I}_{3,0}} = x_{I_{1,0}}x_{I_{2,0}}x_{I_{3,0}}$ by a polynomial $h \in \mathcal{G}_d$ such that $\mathrm{LM}(h) = x_{I_{1,0}}x_{I_{2,0}}$. The polynomial $h \in \mathcal{G}_d$ is defined via the sets $\mathcal{S}_{1,0}$, $\mathcal{S}_{2,0}$ and $\mathcal{M}_0$ as introduced in step 3) of the algorithm, i.e.,

$$h(\mathbf{x}) = f_{d,\left(I_{\mathcal{M}_0}, I_{\mathcal{M}_0^c}\right)}^{\mathcal{S}_0,\mathcal{M}_0} = x_{I_{1,0}}x_{I_{2,0}} - x_{I_{\mathcal{M}_0}}x_{I_{\mathcal{M}_0^c}} \in \mathcal{G}_d.$$

The division of S^0 by h results in

$$S^0 = \mathrm{LC}(S^0)\Big(x_{I_{3,0}} \cdot f^{\mathcal{S}_0,\mathcal{M}_0}_{d,\left(I_{\mathcal{M}_0},I_{\mathcal{M}_0^c}\right)} + \underbrace{x_{I_{\mathcal{M}_0}} x_{I_{\mathcal{M}_0^c}} x_{I_{3,0}} - \mathrm{NLM}(S^0)}_{=S^1}\Big).$$

Note that by construction

$$I_{\mathcal{M}_0}(k) \le I_{\mathcal{M}_0{}^c}(k) \quad \text{for all } k \in [d]\,. \tag{32}$$

If $S^1 \neq 0$, then in the following iteration $i = 1$ we can write $\mathrm{LM}(S^1) = x_{\mathcal{M}_0} x_{\mathcal{M}_0^c} x_{I_{3,0}}$. Due to (32), a pair $I_{1,1}, I_{2,1}$ as in (30) can be either $I_{\mathcal{M}_0}, I_{3,0}$ or $I_{\mathcal{M}_0^c}, I_{3,0}$. Let us assume the former. Then this iteration results in

$$S^1 = \mathrm{LC}(S^1)\Big(x_{I_{3,1}} \cdot f^{\mathcal{S}_1,\mathcal{M}_1}_{d,\left(I_{\mathcal{M}_1},I_{\mathcal{M}_1^c}\right)} + \underbrace{x_{I_{\mathcal{M}_1}} x_{I_{\mathcal{M}_1^c}} x_{I_{\mathcal{M}_0^c}} - \mathrm{NLM}(S^0)}_{=S^2}\Big)$$

with

$$I_{\mathcal{M}_1}(k) \le I_{\mathcal{M}_0^c}(k), I_{\mathcal{M}_1^c}(k) \quad \text{for all } k \in [d]\,.$$

Next, if $S^2 \neq 0$ and $\mathrm{LM}(S^2) = x_{I_{\mathcal{M}_1}} x_{I_{\mathcal{M}_1^c}} x_{I_{\mathcal{M}_0^c}}$ then a pair of indices satisfying (30) must be $I_{\mathcal{M}_0^c} I_{\mathcal{M}_1^c}$ so that the iteration ends up with

$$S^2 = \mathrm{LC}(S^2)\Big(x_{I_{3,2}} \cdot f^{\mathcal{S}_2,\mathcal{M}_2}_{d,\left(I_{\mathcal{M}_2},I_{\mathcal{M}_2^c}\right)} + \underbrace{x_{I_{\mathcal{M}_1}} x_{I_{\mathcal{M}_2^c}} x_{I_{\mathcal{M}_1}} - \mathrm{NLM}(S^0)}_{=S^3}\Big)$$

such that

$$x_{I_{\mathcal{M}_1}}(k) \le x_{I_{\mathcal{M}_2}}(k) \le x_{I_{\mathcal{M}_2^c}}(k) \quad \text{for all } k \in [d]\,.$$

Thus, in iteration $i = 3$ the leading monomial $\mathrm{LM}(S^3)$ must be $\mathrm{NLM}(S^0)$.

A similar analysis can be performed on the monomial $\mathrm{NLM}(S^0)$ and therefore the algorithm stops after at most 6 iterations. The division algorithm results in

$$S(f_1, f_2) = \sum_{i=0}^{p} \left(\prod_{j=0}^{i} \mathrm{LC}(S^j)\right) x_{I_{3,i}} f^{\mathcal{S}_i,\mathcal{M}_i}_{d,\left(I_{\mathcal{M}_i},I_{\mathcal{M}_i^c}\right)},$$

where $f^{\mathcal{S}_i,\mathcal{M}_i}_{d,\left(I_{\mathcal{M}_i},I_{\mathcal{M}_i^c}\right)} = -x_{I_{\mathcal{M}_i}} x_{I_{\mathcal{M}_i^c}} + x_{I_{1,i}} x_{I_{2,i}} \in \mathcal{G}_d$ and $p \le 5$. All the cases that we left out above are treated in a similar way. This shows that $\mathcal{G}_d$ is a Gröbner basis.

In order to show that $\mathcal{G}_d$ is the *reduced* Gröbner basis, first notice that $\mathrm{LC}(g) = 1$ for all $g \in \mathcal{G}_d$. Furthermore, the leading term of any polynomial in $\mathcal{G}_d$ is of degree two. Thus, it is enough to show that for every pair of different polynomials $f^{\mathcal{S}_1,\mathcal{M}_1}_{d,(I_1,\hat{I}_1)}, f^{\mathcal{S}_2,\mathcal{M}_2}_{d,(I_2,\hat{I}_2)} \in \mathcal{G}_d$ it holds that $\mathrm{LM}(f^{\mathcal{S}_1,\mathcal{M}_1}_{d,(I_1,\hat{I}_1)}) \neq \mathrm{LM}(f^{\mathcal{S}_2,\mathcal{M}_2}_{d,(I_2,\hat{I}_2)})$ with $(I_k, \hat{I}_k) \in \mathcal{T}_d^{\mathcal{S}_k}$ for $k = 1, 2$. But this follows from the fact that all elements of $\mathcal{G}_d$ are different as remarked before the statement of the theorem. □

Remark 4. We have concentrated above on the polynomial ideal generated by all second order minors of all matricizations of the tensor. One may also consider a subset of all possible matricizations corresponding to various tensor decompositions and notions of tensor rank. For example, the Tucker(HOSVD)-rank (corresponding to the Tucker or HOSVD decomposition) of a dth order tensor $\mathbf{X}$ is a d-dimensional vector $\mathbf{r}_{HOSVD} = (r_1, r_2, \ldots, r_d)$ such that $r_i = \operatorname{rank}\left(\mathbf{X}^{\{i\}}\right)$ for all $i \in [d]$, see [24]. The unit unfolding-θ_1-norm (defined above for order four tensors) forms the corresponding relaxation of the tensor nuclear norm.

The tensor train (TT) decomposition is another popular approach for tensor computations [41]. The corresponding TT-rank of a dth order tensor $\mathbf{X}$ is a $(d-1)$-dimensional vector $\mathbf{r}_{TT} = (r_1, r_2, \ldots, r_{d-1})$ such that $r_i = \operatorname{rank}\left(\mathbf{X}^{\{1,\ldots,i\}}\right)$, $i \in [d-1]$. By taking into account only minors of order two of the matricizations $\tau \in \{\{1\}, \{1,2\}, \ldots, \{1,2,\ldots,d-1\}\}$, one may introduce a corresponding polynomial ideal and θ_k-norm.

We leave the investigation of such θ-norms to future contributions.

5. Computational Complexity

The computational complexity of the semidefinite programs for computing the θ_1-norm of a tensor or for minimizing the θ_1-norm subject to a linear constraint depends polynomially on the number of variables, i.e., on the size of $\mathcal{B}_{2k}$, and on the dimension of the moment matrix $\mathbf{M}$. We claim that the overall complexity scales polynomially in n, where for simplicity we consider d-th order tensors in $\mathbb{R}^{n\times n\times\cdots\times n}$. Therefore, in contrast to tensor nuclear norm minimization which is NP-hard for $d\geq 3$, tensor recovery via θ_1-norm minimization is tractable.

Indeed, the moment matrix $\mathbf{M}$ is of dimension $(1+n^d)\times(1+n^d)$ (see also (15) for matrices in $\mathbb{R}^{2\times 2}$) and if $a=n^d$ denotes the total number of entries of a tensor $\mathbf{X}\in\mathbb{R}^{n\times\cdots\times n}$, then the number of the variables is at most $\frac{a\cdot(a+1)}{2}\sim\mathcal{O}(a^2)$ which is polynomial in a. (A more precise counting does not give a substantially better estimate.)

Symmetric tensors. We may reduce the complexity of our semidefinite program by reducing to tensors possessing symmetries. Of course, in practice this requires additional information about the tensors to be recovered. For example, let us consider the case of dth order supersymmetric tensors, i.e., tensors $\mathbf{X}\in\mathbb{R}^{n\times n\times\cdots\times n}$ such that $X_{i_1i_2\ldots i_d}=X_{\sigma(i_1)\sigma(i_2)\ldots\sigma(i_d)}$ for all possible permutations $\sigma:\{i_1,i_2,\ldots,i_d\}\to\{i_1,i_2,\ldots,i_d\}$. In this scenario, the semidefinite program for computing the θ_1-norm is of dimension $(a+1)\times(a+1)$, where

$$a=\binom{n+d-1}{d}\leq\left(e\frac{n+d-1}{d}\right)^d=e^d\left(1+\frac{n-1}{d}\right)^d,$$

where this inequality uses the general estimate $\binom{p}{q}\leq(ep/q)^q$, see e.g. [18, Lemma C.5]. The number of variables in the corresponding semidefinite program for computing the θ_1-norm equals the number of monomials $x_I x_{\hat{I}}$ such that $i_1\leq i_2\leq\ldots\leq i_d\leq\hat{i}_1\leq\ldots\leq\hat{i}_d$, excluding the monomial $x_{11\ldots 1}=\mathrm{LM}(g_d)$, which is

$$\binom{n+2d-1}{2d}-1\leq e^{2d}\left(1+\frac{n-1}{2d}\right)^{2d}.$$

We leave it to future investigation to study in the detail the recovery of low rank supersymmetric tensors via θ_k-minimization.

6. Numerical experiments

Let us now empricially study the performance of low rank tensor recovery via θ_1-norm minimization via numerical experiments, where we concentrate on third order tensors. Given measurements $\mathbf{b}=\mathbf{\Phi}(\mathbf{X})$ of a low rank tensor $\mathbf{X}\in\mathbb{R}^{n_1\times n_2\times n_3}$, where $\mathbf{\Phi}:\mathbb{R}^{n_1\times n_2\times n_3}\to\mathbb{R}^m$ is a linear measurement map, we aim at reconstructing $\mathbf{X}$ as the solution of the minimization program

$$\min\|\mathbf{Z}\|_{\theta_1}\quad\text{subject to }\mathbf{\Phi}(\mathbf{Z})=\mathbf{b}.\tag{33}$$

As outlined in Section 2, the θ_1-norm of a tensor $\mathbf{Z}$ can be computed as the minimizer of the semidefinite program

$$\min_{t,\mathbf{y}} t\quad\text{subject to}\quad\mathbf{M}(t,\mathbf{y},\mathbf{Z})\succcurlyeq 0,$$

where $\mathbf{M}(t,\mathbf{y},\mathbf{X})=\mathbf{M}_{\mathcal{B}_1}(t,\mathbf{X},\mathbf{y})$ is the moment matrix of order 1 associated to the ideal J_3, see Theorem 3. This moment matrix for J_3 is explicitly given by

$$\mathbf{M}(t,\mathbf{y},\mathbf{X})=t\mathbf{M}_0+\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3}X_{ijk}\mathbf{M}_{ijk}+\sum_{i=2}^{9}\sum_{j=1}^{|\mathbf{M}^i|}y_\ell\mathbf{M}_j^i,$$

where $\ell=\sum_{k=3}^{i}\left|\mathbf{M}^{(k-1)}\right|+j$. The matrices $\mathbf{M}_0,\mathbf{M}_{ijk}$ and $\mathbf{M}_j^i$ are provided in Table 6. As discussed in Section 2 for the general case, the θ_1-norm minimization problem (33) is then equivalent to the semidefinite program

$$\min_{t,\mathbf{y},\mathbf{Z}} t\quad\text{subject to }\mathbf{M}(t,\mathbf{y},\mathbf{Z})\succeq 0\quad\text{and}\quad\mathbf{\Phi}(\mathbf{Z})=\mathbf{b}.\tag{34}$$

	θ-basis	position (p,q) in the matrix	M_{pq}	Range of $i,\hat{i},j,\hat{j},k,\hat{k}$
$\mathbf{M}_0$	1	$(1,1),(2,2)$	1	
$\mathbf{M}_{ijk}$	x_{ijk}	$(1,f(i,j,k))$	1	$i\in[n_1],j\in[n_2],k\in[n_3]$
$\mathbf{M}^2_{f_2}$	x^2_{ijk}	$(2,2)$	-1	
		$(f(i,j,k),f(i,j,k))$	1	$\{i\in[n_1],j\in[n_2],k\in[n_3]\}$ $\setminus\{i=j=k=1\}$
$\mathbf{M}^3_{f_3}$	$x_{i\hat{j}k}x_{ij\hat{k}}$	$(f(i,j,k),f(i,\hat{j},\hat{k})),(f(i,j,\hat{k}),f(i,\hat{j},k))$	1	$i\in[n_1],j<\hat{j},k<\hat{k}$
$\mathbf{M}^4_{f_4}$	$x_{ijk}x_{\hat{i}\hat{j}\hat{k}}$	$(f(i,j,k),f(\hat{i},\hat{j},\hat{k})),(f(i,\hat{j},k),f(\hat{i},j,\hat{k}))$	1	
		$(f(i,\hat{j},\hat{k}),f(\hat{i},j,k)),(f(i,j,\hat{k}),f(\hat{i},\hat{j},k))$	1	$i<\hat{i},j<\hat{j},k<\hat{k}$
$\mathbf{M}^5_{f_5}$	$x_{ijk}x_{\hat{i}j\hat{k}}$	$(f(i,j,k),f(\hat{i},j,\hat{k})),(f(i,j,\hat{k}),f(\hat{i},j,k))$	1	$i<\hat{i},j\in[n_2],k<\hat{k}$
$\mathbf{M}^6_{f_6}$	$x_{ijk}x_{\hat{i}\hat{j}k}$	$(f(i,j,k),f(\hat{i},\hat{j},k)),(f(i,\hat{j},k),f(\hat{i},j,k))$	1	$i<\hat{i},j<\hat{j},k\in[n_3]$
$\mathbf{M}^7_{f_7}$	$x_{\hat{i}jk}x_{ijk}$	$(f(i,j,k),f(\hat{i},j,k))$	1	$i<\hat{i},j\in[n_2],k\in[n_3]$
$\mathbf{M}^8_{f_8}$	$x_{i\hat{j}k}x_{ijk}$	$(f(i,j,k),f(i,\hat{j},k))$	1	$i\in[n_1],j<\hat{j},k\in[n_3]$
$\mathbf{M}^9_{f_9}$	$x_{ij\hat{k}}x_{ijk}$	$(f(i,j,k),f(i,j,\hat{k}))$	1	$i\in[n_1],j\in[n_2],k<\hat{k}$

TABLE 6. The matrices involved in the definition of the moment matrix $\mathbf{M}(t,\mathbf{X},\mathbf{y})$. Due to symmetry only the upper triangle part of the matrices is specified. The non-specified entries of the matrices $\mathbf{M}\in\mathbb{R}^{(n_1n_2n_3+1)\times(n_1n_2n_3+1)}$ from the first column are equal to zero. The index f_j of $\mathbf{M}^j_{f_j}$ corresponding to the element f_j+J of the θ-basis is specified in the second column. For $\ell\in[9]\setminus\{1\}$, the function f_ℓ denotes an arbitrary but fixed bijection $\{(i,\hat{i},j,\hat{j},k,\hat{k})\}\mapsto\{1,2,\ldots,\left|\mathbf{M}^\ell\right|\}$, where $\mathbf{M}^\ell=\{\mathbf{M}^\ell_{f_\ell}\}$ with $i,\hat{i},j,\hat{j},k,\hat{k}$ in the range of the last column. The function $f:\mathbb{Z}^3\to\mathbb{R}$ is defined as $f(i,j,k)=(i-1)n_2n_3+(j-1)n_3+k+1$.

For our experiments, the linear mapping is defined as $(\mathbf{\Phi}(\mathbf{X}))_k=\langle\mathbf{X},\mathbf{\Phi}_k\rangle$, $k\in[m]$, with independent Gaussian random tensors $\mathbf{\Phi}_k\in\mathbb{R}^{n_1\times n_2\times n_3}$, i.e., all entries of $\mathbf{\Phi}_k$ are independent $\mathcal{N}\left(0,\frac{1}{m}\right)$ random variables. We choose tensors $\mathbf{X}\in\mathbb{R}^{n_1\times n_2\times n_3}$ of rank one as $\mathbf{X}=\mathbf{u}\otimes\mathbf{v}\otimes\mathbf{w}$, where each entry of the vectors $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ is taken independently from the normal distribution $\mathcal{N}(0,1)$. Tensors $\mathbf{X}\in\mathbb{R}^{n_1\times n_2\times n_3}$ of rank two are generated as the sum of two random rank one tensors. With $\mathbf{\Phi}$ and $\mathbf{X}$ given, we compute $\mathbf{b}=\mathbf{\Phi}(\mathbf{X}$, run the semidefinite program (34) and compare its minimizer with the original low rank tensor $\mathbf{X}$. For a given set of parameters, i.e., dimensions n_1,n_2,n_3, number of measurements m and rank r, we repeat this experiment 200 times and record the empirical success rate of recovering the original tensor, where we say that recovery is successful if the elementwise reconstruction error is at most 10^{-6}. We use MATLAB (R2008b) for these numerical experiments, including SeDuMi_1.3 for solving the semidefinite programs.

Table 7 summarizes the results of our numerical tests for cubic and non-cubic tensors of rank one and two and several choices of the dimensions. Here, the number $m_{\max}$ denotes the maximal number of measurements for which not even one out of 200 generated tensors is recovered and $m_{\min}$ denotes the minimal number of measurements for which all 200 tensors are recovered.The fifth column in Table 7 represents the number of independent measurements which are always sufficient for the recovery of a tensor of an arbitrary rank. For illustration, we present the average cpu time (in seconds) for solving the semidefinite programs via SeDuMi_1.3 in the last column. We remark that no attempt of accelerating the optimization algorithm has been made. This task is left for future research.

Except for very small tensor dimensions, we can always recover tensors of rank one or two from a number of measurements which is significantly smaller than the dimension of the corresponding

$n_1 \times n_2 \times n_3$	rank	$m_{\max}$	$m_{\min}$	$n_1n_2n_3$	cpu(sec)
$2\times2\times3$	1	4	12	12	0.1976
$3\times3\times3$	1	6	19	27	0.3705
$3\times4\times5$	1	11	30	60	6.6600
$4\times4\times4$	1	11	32	64	7.2818
$4\times5\times6$	1	18	42	120	129.4804
$5\times5\times5$	1	18	43	125	138.9040
$3\times4\times5$	2	27	56	60	7.5494
$4\times4\times4$	2	26	56	64	8.6525
$4\times5\times6$	2	41	85	120	192.5787

TABLE 7. Numerical results for low rank tensor recovery in $\mathbb{R}^{n_1\times n_2\times n_3}$.

tensor space. Therefore, low rank tensor recovery via θ_1-minimization seems to be a promising approach. Of course, it remains to investigate the recovery performance theoretically.

7. Appendix: Monomial orderings and Gröbner bases

An ordering on the set of monomials $\mathbf{x}^\alpha \in \mathbb{R}[\mathbf{x}]$, $\mathbf{x}^\alpha = x_1^{\alpha_1} \cdot x_2^{\alpha_2} \cdot x_n^{\alpha_n}$, is essential for dealing with polynomial ideals. For instance, it determines an order in a multivariate polynomial division algorithm. Of particular interest is the *graded reverse lexicographic (grevlex) ordering.*

Definition 6. For $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_n)$, $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_n) \in \mathbb{Z}^n_{\geq 0}$, we write $\mathbf{x}^{\boldsymbol{\alpha}} >_{grevlex} \mathbf{x}^{\boldsymbol{\beta}}$ (or $\boldsymbol{\alpha} >_{grevlex} \boldsymbol{\beta}$) if $|\boldsymbol{\alpha}| > |\boldsymbol{\beta}|$ or $|\boldsymbol{\alpha}| = |\boldsymbol{\beta}|$ and the rightmost nonzero entry of $\boldsymbol{\alpha} - \boldsymbol{\beta}$ is negative.

Once a monomial ordering is fixed, the meaning of leading monomial, leading term and leading coefficient of a polynomial (see Section 2) is well-defined. For more information on monomial orderings, we refer the interested reader to [13, 12].

A Gröbner basis is a particular kind of generating set of a polynomial ideal. It was first introduced in 1965 in the Phd thesis of Buchberger [4].

Definition 7 (Gröbner basis)**.** For a fixed monomial order, a basis $\mathcal{G} = \{g_1, g_2, \ldots, g_s\}$ of a polynomial ideal $J \subset \mathbb{R}[\mathbf{x}]$ is a *Gröbner basis* (or standard basis) if for all $f \in \mathbb{R}[\mathbf{x}]$ there exist a **unique** $r \in \mathbb{R}[\mathbf{x}]$ and $g \in J$ such that

$$f = g + r$$

and no monomial of r is divisible by any of the leading monomials in $\mathcal{G}$, i.e., by any of the monomials $\mathrm{LM}(g_1), \mathrm{LM}(g_2), \ldots, \mathrm{LM}(g_s)$.

A Gröbner basis is not unique, but the reduced version defined next is.

Definition 8. The *reduced Gröbner basis* for a polynomial ideal $J \in \mathbb{R}[\mathbf{x}]$ is a Gröbner basis $\mathcal{G} = \{g_1, g_2, \ldots, g_s\}$ for J such that

1) $\mathrm{LC}(g_i) = 1$, for all $i \in [s]$.
2) g_i does not belong to $\langle \mathrm{LT}(\mathcal{G}\backslash\{g_i\})\rangle$ for all $i \in [s]$.

In other words, a Gröbner basis $\mathcal{G}$ is the reduced Gröbner basis if for all $i \in [s]$ the polynomial $g_i \in \mathcal{G}$ is monic (i.e., $\mathrm{LC}(g_i) = 1$) and the leading monomial $\mathrm{LM}(g_i)$ does not divide any monomial of g_j, $j \neq i$.

Many important properties of the ideal and the corresponding algebraic variety can be deduced via its (reduced) Gröbner basis. For example, a polynomial belongs to a given ideal if and only if the unique r from the Definition 7 equals zero. Gröbner bases are also one of the main computational tools in solving systems of polynomial equations [13].

With $\overline{f}^F$ we denote the remainder on division of f by the ordered k-tuple $F = (f_1, f_2, \ldots, f_k)$. If F is a Gröbner basis for an ideal $\langle f_1, f_2, \ldots, f_k\rangle$, then we can regard F as a set without any particular order by Definition 7, or in other words, the result of the division algorithm does not depend on the order of the polynomials. Therefore, $\overline{f}^{\mathcal{G}} = r$ in Definition 7.

The following result follows directly from Definition 7 and the polynomial division algorithm [13].

Corollary 2. *Fix a monomial ordering and let* $\mathcal{G} = \{g_1, g_2, \ldots, g_s\} \subset \mathbb{R}[\mathbf{x}]$ *be a Groebner basis of a polynomial ideal* J*. A polynomial* $f \in \mathbb{R}[\mathbf{x}]$ *is in the ideal* J *if it can be written in the form* $f = a_1 g_1 + a_2 g_2 + \ldots + a_s g_s$*, where* $a_i \in \mathbb{R}[\mathbf{x}]$*, for all* $i \in [s]$*, s.t. whenever* $a_i g_i \neq 0$ *we have*

$$\text{multideg}(f) \geq \text{multideg}(a_i g_i).$$

Definition 9. Fix a monomial order and let $\mathcal{G} = \{g_1, g_2, \ldots, g_s\} \subset \mathbb{R}[\mathbf{x}]$. Given $f \in \mathbb{R}[\mathbf{x}]$, we say that f reduces to zero modulo $\mathcal{G}$ and write

$$f \to_{\mathcal{G}} 0$$

if it can be written in the form $f = a_1 g_1 + a_2 g_2 + \ldots + a_k g_k$ with $a_i \in \mathbb{R}[\mathbf{x}]$ for all $i \in [k]$ s.t. whenever $a_i g_i \neq 0$ we have $\text{multideg}(f) \geq \text{multideg}(a_i g_i)$.

Assume that $\mathcal{G}$ in the above definition is a Gröbner basis of a given ideal J. Then a polynomial f is in the ideal J if and only if f reduces to zero modulo $\mathcal{G}$. In other words, for a Gröbner basis $\mathcal{G}$,

$$f \to_{\mathcal{G}} 0 \quad \text{if and only if} \quad \overline{f}^{\mathcal{G}} = 0.$$

The Gröbner basis of a polynomial ideal always exists and can be computed in a finite number of steps via Buchberger's algorithm [4, 13, 12].

Next we define the S-polynomial of given polynomials f and g which is important for checking whether a given basis of the ideal is a Gröbner basis.

Definition 10. Let $f, g \in \mathbb{R}[\mathbf{x}]$ be a non-zero polynomials.

(1) If $\text{multideg}(f) = \boldsymbol{\alpha}$ and $\text{multideg}(g) = \boldsymbol{\beta}$, then let $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_n)$, where $\gamma_i = \max\{\alpha_i, \beta_i\}$, for every i. We call $\mathbf{x}^{\boldsymbol{\gamma}}$ the least common multiple of $\text{LM}(f)$ and $\text{LM}(g)$ written $\mathbf{x}^{\boldsymbol{\gamma}} = \text{LCM}(\text{LM}(f), \text{LM}(g))$.
(2) The S-polynomial of f and g is the combination

$$S(f, g) = \frac{\mathbf{x}^{\boldsymbol{\gamma}}}{\text{LT}(f)} f - \frac{\mathbf{x}^{\boldsymbol{\gamma}}}{\text{LT}(g)} g.$$

The following theorem gives a criterion for checking whether a given basis of a polynomial ideal is a Gröbner basis.

Theorem 6 (Buchberger's criterion)**.** *A basis* $\mathcal{G} = \{g_1, g_2, \ldots, g_s\}$ *for a polynomial ideal* $J \subset \mathbb{R}[\mathbf{x}]$ *is a Gröbner basis if and only if* $S(g_i, g_j) \to_{\mathcal{G}} 0$ *for all* $i \neq j$*.*

Computing whether $S(g_i, g_j) \to_{\mathcal{G}} 0$ for all possible pairs of polynomials in the basis $\mathcal{G}$ can be a tedious task. The following proposition tells us for which pairs of polynomials this is not needed.

Proposition 1. *Given a finite set* $\mathcal{G} \subset \mathbb{R}[\mathbf{x}]$*, suppose that the leading monomials of* $f, g \in \mathcal{G}$ *are relatively prime, i.e.,*

$$\text{LCM}(\text{LM}(f), \text{LM}(g)) = \text{LM}(f)\,\text{LM}(g),$$

then $S(f, g) \to_{\mathcal{G}} 0$.

Therefore, to prove that the set $\mathcal{G} \subset \mathbb{R}[\mathbf{x}]$ is a Gröbner basis, it is enough to show that $S(g_i, g_j) \to_{\mathcal{G}} 0$ for those $i < j$ where $\text{LM}(g_i)$ and $\text{LM}(g_j)$ are not relatively prime.

Acknowledgements

We would like to thank Bernd Sturmfels, Daniel Plaumann and Shmuel Friedland for helpful discussions and useful inputs to this paper. We would also like to thank James Saunderson and Hamza Fawzi for the deeper insights for the matrix case scenario. We acknowledge funding by the European Research Council through the grant StG 258926.

References

[1] R. Bhatia. *Matrix Analysis.* Graduate Texts in Mathematics. 169. Springer, 1996.
[2] G. Blekherman, P. Parrilo, and R. Thomas. *Semidefinite optimization and convex algebraic geometry.* SIAM, 2013.
[3] S. Boyd and L. Vandenberghe. *Convex Optimization.* Cambridge Univ. Press, 2004.
[4] B. Buchberger. Bruno Buchberger's PhD thesis 1965: An algorithm for finding the basis elements of the residue class ring of a zero dimensional polynomial ideal. *J. Symbolic Comput.*, 41(3-4):475–511, 2006.
[5] E. J. Candès and Y. Plan. Tight oracle bounds for low-rank matrix recovery from a minimal number of random measurements. *IEEE Trans. Inform. Theory*, 57(4):2342–2359, 2011.
[6] E. J. Candès and B. Recht. Exact matrix completion via convex optimization. *Found. Comput. Math.*, 9(6):717–772, 2009.
[7] E. J. Candès, T. Strohmer, and V. Voroninski. PhaseLift: exact and stable signal recovery from magnitude measurements via convex programming. *Comm. Pure Appl. Math.*, 66(8):1241–1274, 2013.
[8] E. J. Candès and T. Tao. The power of matrix completion: near-optimal convex relaxation. *IEEE Trans. Information Theory*, 56(5):2053–2080, 2010.
[9] E. J. Candès, T. Tao, and J. K. Romberg. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2):489–509, 2006.
[10] V. Chandrasekaran, B. Recht, P. Parrilo, and A. Willsky. The convex geometry of linear inverse problems. *Found. Comput. Math.*, 12(6):805–849, 2012.
[11] Y. Chen, S. Bhojanapalli, S. Sanghavi, and R. Ward. Completing any low-rank matrix, provably. *ArXiv:1306.2979*, 2013.
[12] D. Cox, J. Little, and D. O'Shea. *Using algebraic geometry*, volume 185 of *Graduate Texts in Mathematics.* Springer, New York, Second edition, 2005.
[13] D. Cox, J. Little, and D. O'Shea. *Ideals, varieties, and algorithms.* Undergraduate Texts in Mathematics. Springer, New York, Third edition, 2007.
[14] C. Da Silva and F. J. Herrmann. Hierarchical tucker tensor optimization-applications to tensor completion. In *SAMPTA 2013*, pages pp–384. Eurasip.
[15] D. L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.
[16] M. F. Duarte and R. G. Baraniuk. Kronecker compressive sensing. *Image Proc., IEEE Trans. on*, 2011.
[17] M. Fazel. *Matrix rank minimization with applications.* PhD thesis, 2002.
[18] S. Foucart and H. Rauhut. *A Mathematical Introduction to Compressive Sensing.* Applied and Numerical Harmonic Analysis. Birkhäuser, 2013.
[19] S. Friedland and L.-H. Lim. Computational complexity of tensor nuclear norm. *ArXiv e-prints*, oct 2014.
[20] S. Gandy, B. Recht, and I. Yamada. Tensor completion and low-n-rank tensor recovery via convex optimization. *Inverse Problems*, 27(2):19pp, 2011.
[21] J. Gouveia, M. Laurent, P. A. Parrilo, and R. Thomas. A new semidefinite programming hierarchy for cycles in binary matroids and cuts in graphs. *Mathematical Programming*, pages 1–23, 2009.
[22] J. Gouveia, P. Parrilo, and R. Thomas. Theta bodies for polynomial ideals. *SIAM J. Optim.*, 20(4):2097–2118, 2010.
[23] F. Grande and R. Sanyal. Theta rank, levelness, and matroid minors. *arXiv preprint arXiv:1408.1262*, 2014.
[24] L. Grasedyck. Hierarchical singular value decomposition of tensors. *SIAM. J. Matrix Anal. & Appl*, 31:2029, 2010.
[25] L. Grasedyck and W. Hackbusch. An introduction to hierarchical (h-) rank and tt-rank of tensors with examples. *Computational Methods in Applied Mathematics Comput. Methods Appl. Math.*, 11(3):291–304, 2011.
[26] D. Gross. Recovering low-rank matrices from few coefficients in any basis. *IEEE Trans. Inform. Theory*, 57(3):1548–1566, 2011.
[27] D. Gross, Y.-K. Liu, S. FlammiaT., S. Becker, and J. Eisert. Quantum state tomography via compressed sensing. *Phys. Rev. Lett.*, 105:150401, 2010.
[28] W. Hackbusch. *Tensor spaces and numerical tensor calculus.* Springer, 2012.
[29] C. Hillar and L.-H. Lim. Most tensor problems are NP-hard. *J. ACM*, 60(6):45:1–45:39, 2013.
[30] J. Håstad. Tensor rank is NP-complete. *J. Algorithms*, 11(4):644–654, 1990.
[31] B. Huang, C. Mu, D. Goldfarb, and J. Wright. Provable low-rank tensor recovery. *preprint*, 2014.
[32] L. Karlsson, D. Kressner, and A. Uschmajew. Parallel algorithms for tensor completion in the CP format. Sept. 2014.
[33] T. Kolda and B. Bader. Tensor decompositions and applications. *SIAM Rev.*, 51(3):455–500, 2009.
[34] N. Kreimer and M. Sacchi. A tensor higher-order singular value decomposition for prestack seismic data noise reduction and interpolation. *Geophys. J. Internat.*, 77:V113–V122, 2012.
[35] D. Kressner, M. Steinlechner, and B. Vandereycken. Low-rank tensor completion by riemannian optimization. *BIT Numerical Mathematics*, 54(2):447–468, 2014.
[36] R. Kueng, H. Rauhut, and U. Terstiege. Low rank matrix recovery from rank one measurements. *preprint*, 2014.
[37] J. Lasserre. *Moments, positive polynomials and their applications*, volume 1 of *Imperial College Press Optimization Series.* Imperial College Press, London, 2010.

[38] J. Liu, P. Musialski, P. Wonka, and J. Ye. Tensor completion for estimating missing values in visual data. In *ICCV*, 2009.
[39] Y. Liu, F. Shang, W. Fan, J. Cheng, and H. Cheng. Generalized higher-order orthogonal iteration for tensor decomposition and completion. In *Advances in Neural Information Processing Systems*, pages 1763–1771, 2014.
[40] L. Lovász. On the Shannon capacity of a graph. *IEEE Trans. Inform. Theory*, 25(1):1–7, 1979.
[41] I. Oseledets. Tensor-train decomposition. *SIAM J. Sci. Comput.*, 33(5):2295–2317, 2011.
[42] S. Oymak, A. Jalali, M. Fazel, Y. C. Eldar, and B. Hassibi. Simultaneously structured models with application to sparse and low-rank matrices. *IEEE Trans. Inform. Theory*, to appear.
[43] N. Parikh and S. Boyd. Proximal algorithms. *Foundations and Trends in Optimization*, 1(3):123–231, 2013.
[44] H. Rauhut, R. Schneider, and Ž. Stojanac. Tensor tensor recovery via iterative hard thresholding. In *Proc. SampTA 2013*, 2013.
[45] H. Rauhut, R. Schneider, and Ž. Stojanac. Tensor completion in hierarchical tensor representations. In H. Boche, R. Calderbank, G. Kutyniok, and J. Vybiral, editors, *Compressed sensing and its applications*. Springer, 2015.
[46] H. Rauhut, R. Schneider, and Ž. Stojanac. Low rank tensor tensor recovery via iterative hard thresholding. in preparation.
[47] B. Recht, M. Fazel, and P. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Rev.*, 52(3):471–501, 2010.
[48] B. Romera Paredes, H. Aung, N. Bianchi Berthouze, and M. Pontil. Multilinear multitask learning. *J. Mach. Learn. Res.*, 28(3):1444–1452, 2013.
[49] K. Toh and S. Yun. An accelerated proximal gradient algorithm for nuclear norm regularized least squares problems. *Pac. J. Optim.*, 6:615–640, 2010.
[50] R. Tomioka, K. Hayashi, and H. Kashima. Estimation of low-rank tensors via convex optimization. *Preprint*, 2010. ArXiv:1010.0789.
[51] M. Yuan and C.-H. Zhang. On tensor completion via nuclear norm minimization. *preprint*, 2014. ArXiv:1405.1773.

RWTH Aachen University, Lehrstuhl C für Mathematik (Analysis), Pontdriesch 10, 52062 Aachen Germany
E-mail address: rauhut@mathc.rwth-aachen.de

RWTH Aachen University, Lehrstuhl C für Mathematik (Analysis), Pontdriesch 10, 52062 Aachen Germany and University of Bonn, Hausdorff Center for Mathematics, Endenicher Allee 62, 53115 Bonn, Germany
E-mail address: zeljka.stojanac@hcm.uni-bonn.de