TENSOR THETA NORMS AND LOW RANK RECOVERY

HOLGER RAUHUT AND ŽELIKA 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-understood 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 in general 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 \( \theta_k \)-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. The sequence of the corresponding unit-\( \theta_k \)-norm balls converges asymptotically to the unit tensor nuclear norm ball. By providing the Gröbner basis for the ideals, we explicitly give semidefinite programs for the computation of the \( \theta_k \)-norm and for the minimization of the \( \theta_k \)-norm under an affine constraint. Finally, numerical experiments for order-three tensor recovery via \( \theta_1 \)-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 \( \ell_1 \)-minimization \cite{Donoho2006, Candes2006, Candes2007}. 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 \cite{Rauhut2009, Lin2010}. 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 \cite{Wright2009, Gillis2010, Fornasier2010, Candes2011, Krahmer2011, Wang2011, Zhou2011, Zhou2011}.

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 \cite{Kang2010}, reflectance data...
recovery [45] (e.g. for use in photo-realistic raytracers), machine learning [55], and seismic data processing [40]. Several approaches have already been introduced [24, 39, 45, 51, 52], but unfortunately, so far, for none of them a completely satisfactory theory is available. Either the method is not tractable [62], or no (complete) rigorous recovery results quantifying the minimal number of measurements are available [16, 24, 39, 41, 45, 51, 52], or the available bounds are highly nonoptimal [20, 38, 46]. For instance, the computation (and therefore, also the minimization) of the tensor nuclear norm [18, 56, 60] for higher order tensors is in general NP-hard [23] – nevertheless, some recovery results for tensor completion via nuclear norm minimization are available in [62]. Moreover, versions of iterative hard thresholding for various tensor formats have been introduced [51, 52]. This approach leads to a computationally tractable algorithm, which empirically works well. However, only a partial analysis based on the tensor 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 tensor restricted isometry property (TRIP) has been analyzed for certain random measurement maps [51–53]. These near optimal bounds on the number of measurements ensuring the TRIP, however, provide only a hint on how many measurements are required because the link between the TRIP and recovery is so far only partial [52, 53].

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 theta bodies, a concept from computational algebraic geometry [2, 26, 47] which is similar to the better known Lasserre relaxations [44]. We arrive at a whole family of convex bodies (indexed by a polynomial degree), which form convex relaxations of the unit nuclear norm ball. The resulting norms are called theta norms. The corresponding unit norm balls are nested and contain the unit nuclear norm ball. Even more, the sequence of the unit-$\theta_k$-norm balls converges asymptotically to the unit tensor nuclear norm ball. They can be computed by semidefinite optimization, and also the minimization of the $\theta_k$-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 unit Frobenius norm. 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 semi-definite programs explicitly and present numerical experiments which show that $\theta_1$-norm minimization successfully recovers tensors of low rank from few random linear measurements. Unfortunately, a rigorous theoretical analysis of the recovery performance of $\theta_k$-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 $X \in \mathbb{R}^{n_1 \times n_2}$ of rank at most $r \ll \min\{n_1, n_2\}$, and suppose we are given linear measurements $y = A(X)$, where $A : \mathbb{R}^{n_1 \times n_2} \rightarrow \mathbb{R}^m$ is a linear map with $m \ll n_1n_2$. Reconstructing $X$ from $y$ amounts to solving an underdetermined linear system. Unfortunately, the rank minimization problem of computing the minimizer of

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

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

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

(1)

has been suggested [21, 54], where the nuclear norm $\|Z\|_* = \sum_j \sigma_j(Z)$ is the sum of the singular values of $Z$. This problem can be solved efficiently by various methods [3]. For instance, it can be reformulated as a semi-definite program [21], but splitting methods may be more efficient [50, 58].

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

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

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

$$m \geq Crn, \quad n = \max\{n_1, n_2\},$$

(2)

where the constants $c, C > 0$ are universal [3] [12]. 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 [7] [9] [13] [30], where the measurements are randomly chosen entries of the matrix $X$. Measurements taken as Frobenius inner products with rank-one matrices are studied in [42], and arise in the phase retrieval problem as special case [8]. 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 $X \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ 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 [29] [35] [36] [43]. We will mainly work with the canonical rank or CP-rank in the
following. A $d$th-order tensor $X \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ is of rank one if there exist vectors $u^1 \in \mathbb{R}^{n_1}$, $u^2 \in \mathbb{R}^{n_2}$, \ldots, $u^d \in \mathbb{R}^{n_d}$ such that $X = u^1 \otimes u^2 \otimes \cdots \otimes u^d$ or elementwise $X_{i_1, i_2, \ldots, i_d} = u_{i_1}^1 u_{i_2}^2 \cdots u_{i_d}^d$.

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

Given a linear measurement map $A : \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d} \to \mathbb{R}^m$ (which can be represented as a $(d+1)$th-order tensor), our aim is to recover a tensor $X \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ from $y = A(X)$ 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,

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

where the nuclear norm is defined as

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

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 [23, 34, 37]. Let us nevertheless mention that some theoretical results for tensor recovery via nuclear norm minimization are contained in [62].

We remark that, unlike in the matrix scenario, the tensor rank and consequently the tensor nuclear norm are dependent on the choice of base field, see for example [4, 17, 23]. In other words, the rank (and the nuclear norm) of a given tensor with real entries depends on whether we regard it as a real tensor or as a complex tensor.

In this paper, we focus only on tensors with real-valued entries, i.e., we work over the field $\mathbb{R}$.

The aim of this article is to introduce relaxations of the tensor nuclear norm, based on 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 $X \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ into a low rank matrix $\tilde{X} \in \mathbb{R}^{n_1 \times n_2}$ and simply apply concepts from matrix recovery. However, the bound (2) on the required number of measurements then reads

$$m \geq Crn^2.$$  \hspace{1cm} (3)

Moreover, it has been suggested in [24, 45, 59] 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 [19], that it cannot work with less measurements than stated by the estimate in (3). This is essentially due to the fact that the tensor structure is not represented. That is, instead of solving a tensor nuclear norm minimization problem under the assumption that the tensor is of low rank, the matrix nuclear norm minimization problem is being solved under the assumption that a particular matricization of a tensor is of low rank.

Bounds for a version of the restricted isometry property for certain tensor formats in [53] suggest that

$$m \geq Cr^2n$$

measurements should be sufficient when working directly with the tensor structure – precisely, this bound uses the tensor train format [48]. (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 [53].) 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.

Unlike in the previously mentioned contributions, we consider the canonical tensor rank and the corresponding tensor nuclear norm, which respects the tensor structure. Even more, it is expected that the bound on the minimal number of measurements needed for low rank tensor recovery via tensor nuclear norm minimization is optimal, see also [62], where tensor completion via tensor nuclear norm minimization has been considered. Unfortunately, it is in general NP-hard to solve this optimization problem (since it is NP-hard to compute the tensor nuclear norm). To overcome this difficulty, in this paper, we provide the tensor \( \theta_k \)-norms – the new tensor norms which can be computed via semidefinite programming. These norms are tightly related to the tensor nuclear norm. That is, the unit \( \theta_k \)-norm balls (which are defined for \( k \in \mathbb{N} \)) satisfy

\[
\{ X : \|X\|_{\theta_k} \leq 1 \} \supseteq \cdots \supseteq \{ X : \|X\|_{\theta_{k+1}} \leq 1 \} \supseteq \{ X : \|X\|_{\theta_{k+2}} \leq 1 \} \supseteq \cdots \supseteq \{ X : \|X\|_* \leq 1 \}.
\]

In particular, we show that in the matrix scenario all \( \theta_k \)-norms coincide with the matrix nuclear norm. In case of order-\( d \) tensors (\( d \geq 3 \)), we prove that the sequence of the unit-\( \theta_k \)-norm balls converges asymptotically to the unit tensor nuclear norm ball. Next, we provide numerical experiments on low rank tensor recovery via \( \theta_1 \)-norm minimization. We provide numerical experiments for \( \theta_1 \)-minimization that indicate that this is a very promising approach for low rank tensor recovery. However, we note that standard solvers for semidefinite programs only allow us to test our method on small to moderate size problems. Nevertheless, it is likely that specialized efficient algorithms can be developed. Indeed, recall that \( \theta_k \)-norms all coincide with the matrix nuclear norm and the state-of-the-art algorithms allow us computing the nuclear norm of matrices of large dimensions. This suggests the possibility that new algorithms could be developed which would allow us to apply our method on larger tensors. Thus, this paper presents the first step in a new convex optimization approach to low rank tensor recovery.

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 \( S \) is denoted by \(|S|\).

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

\[
\{ \det(A_{I,J}) : I \subset [m], J \subset [n], |I| = |J| = k \}.
\]

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

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

where the \( \sigma_i \) list the singular values of \( X \). The nuclear norm is given by \( \|X\|_* = \sum_{i=1}^{\min\{m,n\}} \sigma_i \). It is well-known that its unit ball is the convex hull of all rank-one matrices of unit Frobenius norm.
The vectorization of a tensor $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ is denoted by $\text{vec}(\mathbf{X}) \in \mathbb{R}^{n_1 n_2 \cdots n_d}$. The ordering of the elements in $\text{vec}(\mathbf{X})$ 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 $d$th-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 $d$th-order tensor $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ and an ordered subset $S \subseteq [d]$, an $S$-matricization $\mathbf{X}^S \in \mathbb{R}^{\prod_{k \in S} n_k \times \prod_{\ell \in S^c} n_{\ell}}$ is defined as

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

i.e., the indexes in the set $S$ define the rows of a matrix and the indexes in the set $S^c = [d] \backslash S$ define the columns. For a singelton set $S = \{i\}$, for $i \in [d]$, we call the $S$-matricization the $i$-th unfolding. Notice that every $S$-matricization of a rank-one tensor is a rank-one matrix. Conversely, if every $S$-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 $d$th-order tensor $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$, we write $\mathbf{X}(\cdot, \ldots, k)$ for the $(d-1)$-order subtensor in $\mathbb{R}^{n_1 \times \cdots \times n_{d-1}}$ obtained by fixing the last index $\alpha_d$ to $k$. For simplicity, the subscripts $\alpha_1 \alpha_2 \cdots \alpha_d$ and $\beta_1 \beta_2 \cdots \beta_d$ will often be denoted by $\alpha$ and $\beta$, respectively. In particular, instead of writing $x_{\alpha_1 \alpha_2 \cdots \alpha_d \beta_1 \beta_2 \cdots \beta_d}$, we often just write $x_\alpha x_\beta$. Below, we will use the grevlex ordering of monomials indexed by subscripts $\alpha$, which in particular requires to define an ordering for such subscripts. We make the agreement that $x_{11 \ldots 11} > x_{11 \ldots 12} > \cdots > x_{11 \ldots 1n_d} > x_{111 \ldots 21} > \cdots > x_{n_1 n_2 \ldots n_d}$.

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 unit Frobenius norm matrices. We discuss the corresponding $\theta_k$-norms and show that they all coincide with the matrix nuclear norm. The case of $2 \times 2$-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 $\theta_k$-norms. We additionally show that considering matricizations corresponding to the TT-format will lead to the same polynomial ideal and thus to the same $\theta_k$-norms. The general $d$th-order case is discussed at the end of Section 4. Here, we define the polynomial ideal $J_d$ which corresponds to the set of all possible matricizations of the tensor. We show that a certain set of order-two minors forms the reduced Gröbner basis for this ideal, which is key for defining the $\theta_k$-norms. We additionally show that polynomial ideals corresponding to different tensor formats (such as TT format or Tucker/HOSVD format) coincide with the ideal $J_d$ and consequently, they lead to the same $\theta_k$-norms. In Section 5 we discuss the convergence of the sequence of the unit-$\theta_k$-norm balls to the unit tensor nuclear norm ball. Section 6 briefly discusses the polynomial runtime of the algorithms for computing and minimizing the $\theta_k$-norms showing that our approach is tractable. Numerical experiments for low rank recovery of third-order tensors are presented in Section 7 which show that our approach successfully recovers a low rank tensor from incomplete Gaussian random measurements. Appendix A discusses some background from computer algebra (monomial orderings and Gröbner bases) that is required throughout the main body of the article.
We say that an ideal \( \mathcal{J} \) is \( \nu \)-exact if \( \{ x \in \mathbb{R}^n : f(x) = 0 \} \) is finite for all \( f \in \mathcal{J} \). Throughout the paper, \( \mathbb{R}[x] \) denotes the ring of polynomials in \( n \) variables. Let \( \mathcal{J} \subseteq \mathbb{R}[x] \) be a polynomial ideal. Its real algebraic variety is the set of all points in \( x \in \mathbb{R}^n \) where all polynomials in the ideal vanish, i.e.,

\[
\nu_{\mathbb{R}}(\mathcal{J}) = \{ x \in \mathbb{R}^n : f(x) = 0 \text{ for all } f \in \mathcal{J} \}.
\]

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

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

Its real algebraic variety is the set

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

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

**Definition 2.1.** Let \( \mathcal{J} \) be an ideal in \( \mathbb{R}[x] \). A polynomial \( f \in \mathbb{R}[x] \) is \( k \)-sos mod \( \mathcal{J} \) if there exists a finite set of polynomials \( h_1, h_2, \ldots, h_t \in \mathbb{R}[x]_k \) such that \( f = \sum_{j=1}^t h_j^2 \) mod \( \mathcal{J} \), i.e., if \( f - \sum_{j=1}^t h_j^2 \in \mathcal{J} \).

A special case of theta bodies was first introduced by Lovász in [47] and in full generality they appeared in [25]. Later, they have been analyzed in [25, 27]. The definitions and theorems in the remainder of the section are taken from [25].

**Definition 2.2 (Theta body).** Let \( \mathcal{J} \subseteq \mathbb{R}[x] \) be an ideal. For a positive integer \( k \), the \( k \)-th \emph{theta body} of \( \mathcal{J} \) is defined as

\[
TH_k(\mathcal{J}) := \{ x \in \mathbb{R}^n : f(x) \geq 0 \text{ for every linear } f \text{ that is } k\text{-sos mod } \mathcal{J} \}.
\]

We say that an ideal \( \mathcal{J} \subseteq \mathbb{R}[x] \) is \( TH_k \)-exact if \( TH_k(\mathcal{J}) \) equals \( \nu_{\mathbb{R}}(\mathcal{J}) \), the closure of the convex hull of \( \nu_{\mathbb{R}}(\mathcal{J}) \).

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

\[
\text{TH}_1(\mathcal{J}) \supseteq \text{TH}_2(\mathcal{J}) \supseteq \cdots \supseteq \text{conv}(\nu_{\mathbb{R}}(\mathcal{J})).
\]

The theta-body sequence of \( \mathcal{J} \) can converge (finitely or asymptotically), if at all, only to \( \text{conv}(\nu_{\mathbb{R}}(\mathcal{J})) \). More on guarantees on convergence can be found in [25, 27]. 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 \( \mathcal{J} \) 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 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 \(\mathbb{R}[x^i] = \mathbb{R}[x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n]\). Therefore, \(\mathbb{R}[x^i]/J \cong \mathbb{R}[x^i]_1 \) and \(\{1 + J, x_1 + J, \ldots, x_n + J\}\) can be completed to a basis \(B\) of \(\mathbb{R}[x]/J\). 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 \(B = \{f_i + J\}\) of \(\mathbb{R}[x]/J\) 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 \(B = \{f_i + J\}\) is ordered so that \(f_{i+1} > f_i\), where \(>\) is a fixed monomial ordering. Further, we define the set \(B_k\)

\[B_k := \{f + J \in B : \deg (f + J) \leq k\}\.

**Definition 2.3** (Theta basis). Let \(J \subseteq \mathbb{R}[x]\) be an ideal. A basis \(B = \{f_0 + J, f_1 + J, \ldots\}\) of \(\mathbb{R}[x]/J\) is a \(\theta\)-basis if it has the following properties

1) \(B_1 = \{1 + J, x_1 + J, \ldots, x_n + J\}\),
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 \(B_{2k}\).

As in [2, 26] we consider only monomial bases \(B\) of \(\mathbb{R}[x]/J\), i.e., bases \(B\) such that \(f_i\) is a monomial, for all \(f_i + J \in \mathcal{B}\).

For determining a \(\theta\)-basis, we first need to compute the reduced Gröbner basis \(\mathcal{G}\) of the ideal \(J\), see Definitions A.2 and A.3. The set \(B\) 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 A.1) or simply grevlex ordering, although also the graded lexicographic ordering would be appropriate.

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

\[J_{\text{initial}} = \left\{ \deg(f_i) \right\}_{f_i \in J} = \left\{ \deg(g_i) \right\}_{g_i \in [s]} ,
\]

where \(\mathcal{G} = \langle g_1, g_2, \ldots, g_k \rangle\) is the reduced Gröbner basis of the ideal \(J\). In other words, a set \(B = \{f_0 + J, f_1 + J, \ldots\}\) will be a \(\theta\)-basis of \(\mathbb{R}[x]/J\) 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 \(\{\deg(g_i) : i \in [s]\}\).

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

The theta bodies can be characterized via the combinatorial moment matrix as stated in the next result from [26], which will be the basis for computing and minimization the new tensor norm introduced below via semidefinite programming.

**Definition 2.4.** Let \(J, B\) and \(\{\lambda_{i,j}\}\) be as above. Let \(y\) be a real vector indexed by \(B_{2k}\) with \(y_0 = 1\), where \(y_0\) is the first entry of \(y\), indexed by the basis element \(1 + J\). The \(k\)-th combinatorial moment matrix \(M_{B_k}(y)\) of \(J\) is the real matrix indexed by \(B_k\) whose \((i,j)\)-entry is \(M_{B_k}(y)_{i,j} = \sum_{f_{i,j} + J \in B_{2k}} \lambda_{i,j} y_{i,j}\).
Theorem 2.5. The k-th theta body of J, \( \Theta_k(J) \), is the closure of
\[
Q_{B_k}(J) = \pi_{\mathbb{R}^n} \left\{ y \in \mathbb{R}^{B_k} : M_{B_k}(y) \succeq 0, y_0 = 1 \right\},
\]
where \( \pi_{\mathbb{R}^n} \) denotes the projection onto the variables \( y_1 = y_{x_1 + J}, \ldots, y_n = y_{x_n + J} \).

Algorithm 1 shows a step-by-step procedure for computing \( \Theta_k(J) \).

**Algorithm 1** Algorithm for computing \( \Theta_k(J) \)

**Input:** An ideal \( J \in \mathbb{R}[x] = \mathbb{R}[x_1, x_2, \ldots, x_n] \).

**Compute** the reduced Gröbner basis for the ideal \( J \)

**Compute** a \( \theta \)-basis \( B = B_1 \cup B_2 \cup \ldots = \{ f_0 + J, f_1 + J, \ldots \} \) of \( \mathbb{R}[x]/J \) (see Definition 2.3)

**Compute** the combinatorial moment matrix \( M_{B_k}(y) \):

1. \( [x]_{B_k} = \{ \text{all elements of } B_k \text{ in order} \} \)
2. \( (X_{B_k})_{i,j} = \left( [x]_{B_k} [x]_{B_k}^T \right)_{i,j} = f_i f_j + J = \sum_{f_i + J \in B_k} \lambda_{i,j}^k (f_i + J) \)
3. \( [M_{B_k}(y)]_{i,j} = \sum_{f_i + J \in B_k} \lambda_{i,j}^k y_l \)

**Output:** \( \Theta_k(J) \) is the closure of \( Q_{B_k}(J) = \pi_{\mathbb{R}^n} \left\{ y \in \mathbb{R}^{B_k} : M_{B_k}(y) \succeq 0, y_0 = 1 \right\} \).

### 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 \( \theta_k \)-norm ball with the property
\[
\|X\|_{\theta_k} \leq \|X\|_{\theta_{k+1}} \quad \text{for all } X \in \mathbb{R}^{m \times n} \text{ and all } k \in \mathbb{N}.
\]

However, we will show that all these \( \theta_k \)-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, unit Frobenius norm matrices
\[
\nu_k(J) = \left\{ X \in \mathbb{R}^{m \times n} : \|X\|_F = 1, \ \text{rank} \ (X) = 1 \right\}.
\] (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 \( \mathbb{R}[x] = \mathbb{R}[x_{11}, x_{12}, \ldots, x_{mn}] \)
\[
g(x) = \sum_{i=1}^{m} \sum_{j=1}^{n} x_{ij}^2 - 1 \quad \text{and} \quad f_{ijkl}(x) = x_{ik} x_{kj} - x_{ij} x_{kl}
\]
for \( 1 \leq i < k \leq m, 1 \leq j < l \leq n \). (6)

**Lemma 3.1.** Let \( X \in \mathbb{R}^{m \times n} \setminus \{0\} \). Then \( X \) is a rank-one, unit Frobenius norm matrix if and only if
\[
X \in \mathcal{R} := \{ X : g(X) = 0 \text{ and } f_{ijkl}(X) = 0 \text{ for all } i < k, j < l \}.
\] (7)
Proof. If \( X \in \mathbb{R}^{m \times n} \) is a rank-one matrix with \( \|X\|_F = 1 \), then by definition there exist two vectors \( u \in \mathbb{R}^m \) and \( v \in \mathbb{R}^n \) such that \( X_{ij} = u_i v_j \) for all \( i \in [m], j \in [n] \) and \( \|u\|_2 = \|v\|_2 = 1 \). Thus
\[
X_{ij}X_{kl} - X_{il}X_{kj} = u_i v_j u_k v_l - u_i v_j u_k v_j = 0
\]
and
\[
\sum_{i=1}^m \sum_{j=1}^n X^2_{ij} = \sum_{i=1}^m u_i^2 \sum_{j=1}^n v_j^2 = 1.
\]
For the converse, let \( X_i \) represent the \( i \)-th column of a matrix \( X \in \mathcal{R} \). Then, for all \( j, l \in [n] \) with \( j < l \), it holds
\[
X_{ml} \cdot X_j - X_{mj} \cdot X_l = \begin{bmatrix} X_{1j}X_{ml} - X_{1l}X_{mj} \\ X_{2j}X_{ml} - X_{2l}X_{mj} \\ \vdots \\ X_{mj}X_{ml} - X_{mj}X_{ml} \end{bmatrix} = 0,
\]
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 \( X \) span a space of dimension one, i.e., the matrix \( X \) is a rank-one matrix. From \( \sum_{i=1}^m \sum_{j=1}^n X^2_{ij} - 1 = 0 \) it follows that the matrix \( X \) is normalized, i.e., \( \|X\|_F = 1 \).

It follows from Lemma 3.1 that the set of rank-one, unit Frobenius norm matrices coincides with the algebraic variety \( \nu_{\mathcal{R}}(J_{M_{mn}}) \) for the ideal \( J_{M_{mn}} \) generated by the polynomials \( g \) and \( f_{ijkl} \), i.e.,
\[
J_{M_{mn}} = \langle \mathcal{G}_{M_{mn}} \rangle \quad \text{with}
\]
\[
\mathcal{G}_{M_{mn}} = \{g(x)\} \cup \{f_{ijkl}(x) : 1 \leq i < k \leq m, 1 \leq j < l \leq n\}. \tag{8}
\]
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,
\[
\text{conv} (\nu_{\mathcal{R}}(J_{M_{mn}})) \subseteq \cdots \subseteq \text{TH}_{k+1}(J_{M_{mn}}) \subseteq \text{TH}_{k}(J_{M_{mn}}) \subseteq \cdots \subseteq \text{TH}_{1}(J_{M_{mn}}).
\]
Therefore, the theta bodies form closed, convex hierarchical relaxations of the matrix nuclear norm ball. In addition, the theta body \( \text{TH}_{k}(J_{M_{mn}}) \) is symmetric, \( \text{TH}_{k}(J_{M_{mn}}) = -\text{TH}_{k}(J_{M_{mn}}) \). Therefore, it defines a unit ball of a norm that we call the \( \theta_k \)-norm.

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

**Lemma 3.2.** 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 A.8 in the appendix, we only need to check whether the \( S \)-polynomial, see Definition A.6, satisfies \( S(p, q) \rightarrow \mathcal{G}_{M_{mn}} 0 \) for all \( p, q \in \mathcal{G}_{M_{mn}} \) whenever the leading monomials \( \text{LM}(p) \) and \( \text{LM}(q) \) are not relatively prime. Here, \( S(p, q) \rightarrow \mathcal{G}_{M_{mn}} 0 \) means that \( S(p, q) \) reduces to 0 modulo \( \mathcal{G}_{M_{mn}} \), see Definition A.5.

Notice that \( \text{LM}(g) = x_1^2 \) and \( \text{LM}(f_{ijkl}) = x_{il}x_{kj} \) 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 \( S(f_{ijkl}, f_{ijkl}') \rightarrow \mathcal{G}_{M_{mn}} 0 \) whenever the leading monomials \( \text{LM}(f_{ijkl}) \) and \( \text{LM}(f_{ijkl}') \) are not relatively prime. First we consider
\[
f_{ijkl}(x) = x_{il}x_{kj} - x_{ij}x_{kl} \quad \text{and} \quad f_{ijkl}'(x) = x_{il}x_{kj} - x_{ij}x_{kl}
\]
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
\[
S(f_{ijk}, f_{ijk'}) = x_{ij} f_{ijk}(x) - x_{k} f_{ijk'}(x) = -x_{ij} x_{k} f_{ijk'}(x) + x_{ij} x_{k} x_{k'}
\]
so that $S(f_{ijk}, f_{ijk'}) \rightarrow \mathcal{G}_{mn}$ 0. The remaining cases are treated with similar arguments.

In order to show that $\mathcal{G}_{mn}$ is a reduced Gröbner basis (see Definition 3.3), we first notice that LC($f$) = 1 for all $f \in \mathcal{G}_{mn}$. In addition, the leading monomial of $f \in \mathcal{G}_{mn}$ is always of degree two and there are no two different polynomials $f_1, f_2 \in \mathcal{G}_{mn}$ such that $LM(f_1) = LM(f_2)$. Therefore, $\mathcal{G}_{mn}$ is the reduced Gröbner basis of the ideal $J_{mn}$ with respect to the grevlex order.

The Gröbner basis $\mathcal{G}_{mn}$ of $J_{mn}$ ($\mathcal{G}_{mn}$) yields the $\theta$-basis of $\mathbb{R}[x]/J_{mn}$.

For the sake of simplicity, we only provide its elements up to degree two,
\[
\begin{align*}
B_1 &= \{1 + J_{mn}, x_{11} + J_{mn}, x_{12} + J_{mn}, \ldots, x_{mn} + J_{mn} \} \\
B_2 &= B_1 \cup \{x_i x_k J_{mn} : (i, j, k, l) \in \mathcal{S}_{B_2} \},
\end{align*}
\]
where $\mathcal{S}_{B_2} = \{(i, j, k, l) : 1 \leq i \leq k \leq m, 1 \leq j \leq l \leq n \} \setminus \{(1, 1, 1, 1)\}$. Given the $\theta$-basis, the theta body $TH_k(J_{mn})$ is well-defined. We formally introduce an associated norm next.

**Definition 3.3.** The matrix $\theta_k$-norm, denoted by $\| \cdot \|_{\theta_k}$, is the norm induced by the $k$-theta body $TH_k(J_{mn})$, i.e.,
\[
\|X\|_{\theta_k} = \inf \{r : X \in r \cdot \Theta J_{mn} \}.
\]

The $\theta_k$-norm can be computed with the help of Theorem 2.15, i.e., as
\[
\|X\|_{\theta_k} = \min_{t \in \mathbb{R}, y \in \mathbb{R}^{s_k}} t \quad \text{subject to} \quad \mathcal{M}_{B_1}[y] \succeq 0, y_0 = t, \mathcal{M}_{B_1}[y] = X.
\]

The last constraint might require some explanation. The vector $y_{B_1}$ denotes the restriction of $y$ to the indices in $B_1$, where the latter can be identified with the set $[m] \times [n]$ indexing the matrix entries. Therefore, $y_{B_1} = X$ 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 $\theta_2$-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 (21).

Notice that the number of elements in $B_1$ is $mn + 1$, and in $B_2 \setminus B_1$ is $\frac{m(m+1)}{2} \cdot \frac{n(n+1)}{2} - 1 \sim \frac{(mn)^2}{2}$, i.e., the number of elements of the $\theta$-basis restricted to the degree 2 scales polynomially in the total number of matrix entries $mn$. Therefore, the computational complexity of the SDP in (21) is polynomial in $mn$.

We will show next that the theta body $TH_k(J)$ and hence, all $TH_k(J)$ 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 3.4.** Let $O_c$ ($O_r$) denote the set of all matrices $M \in \mathbb{R}^{n \times m}$ with orthonormal columns (rows), i.e., $O_c = \{M \in \mathbb{R}^{n \times m} : M^T M = I_n \}$ and $O_r = \{M \in \mathbb{R}^{n \times m} : MM^T = I_n \}$. Then
\[
\{X \in \mathbb{R}^{m \times n} : \|X\|_* \leq 1\} = \{X \in \mathbb{R}^{m \times n} : \text{tr}(MX) \leq 1, \text{ for all } M \in O_c \cup O_r\}.
\]

**Remark 1.** Notice that $O_c = \emptyset$ for $m > n$ and $O_r = \emptyset$ for $m < n.$
In other words, we start by fixing $M-1 = X$. Let $X \in \mathbb{R}^{m \times n}$ such that $\|X\|_* \leq 1$ and let $X = U \Sigma V^T$ be its singular value decomposition. For $M \in \mathcal{O}_r$, the spectral norm satisfies $|M| \leq 1$ and therefore, using that the nuclear norm is the dual of the spectral norm, see e.g. [3, p. 96],

$$\text{tr}(MX) \leq \|M\| \cdot \|X\|_* \leq \|X\|_* \leq 1.$$ 

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

$$1 \geq \text{tr}(MX) = \text{tr}(V U^T U \Sigma V^T) = \text{tr}(\Sigma) = \|X\|_*.$$ 

This completes the proof.

Next, using Lemma 3.4, we show that the theta body $TH_1(J)$ equals the nuclear norm ball. This result is related to Theorem 4.4 in [27].

**Theorem 3.5.** The polynomial ideal $J_{M,n}$ defined in (8) is $TH_1$-exact, i.e.,

$$TH_1(J_{M,n}) = \text{conv} \{ x : g(x) = 0, f_{ijkl}(x) = 0 \text{ for all } i < k, j < l \}.$$ 

In other words,

$$\{ X \in \mathbb{R}^{m \times n} : X \in TH_1(J_{M,n}) \} = \{ X \in \mathbb{R}^{m \times n} : \|X\|_* \leq 1 \}.$$ 

**Proof.** By definition of $TH_1(J_{M,n})$, it is enough to show that the boundary of the unit nuclear norm can be written as 1-sos mod $J_{M,n}$, which by Lemma 3.4 means that the polynomial $1 - \sum_{i=1}^m \sum_{j=1}^n x_{ij} M_{ji}$ is 1-sos mod $J_{M,n}$ for all $M \in \mathcal{O}_c \cup \mathcal{O}_r$.

We start by fixing $M = \begin{pmatrix} I_m & 0 \\ 0 & 0 \end{pmatrix}$ in case $m \leq n$ and $M = (I_n \ 0)$ in case $m > n$, where $I_k \in \mathbb{R}^{k \times k}$ is the identity matrix. For this choice of $M$, we need to show that $1 - \sum_{i=1}^\ell x_{ii}$ is 1-sos mod $J_{M,n}$, where $\ell = \min \{m, n\}$. Note that

$$1 - \sum_{i=1}^\ell x_{ii} = \frac{1}{2} \left[ \left( 1 - \sum_{i=1}^\ell x_{ii} \right)^2 - \sum_{i<j}^\ell (x_{ii} x_{jj} - x_{ij} x_{ji})^2 \right],$$

since

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

$$1 - \sum_{i=1}^m \sum_{j=1}^n x_{ij}^2 + \sum_{i=1}^m \sum_{j=m+1}^n x_{ij}^2 + \sum_{i=m+1}^n \sum_{j=1}^n x_{ij}^2 = 1 - \sum_{i=1}^\ell \sum_{j=1}^\ell x_{ij}^2$$

$$= 1 - \sum_{i<j}^\ell (x_{ij}^2 + x_{ji}^2) - \sum_{i=1}^\ell x_{ii}^2.$$
Moreover, the following polynomial ideal was suggested

\[ J = \left\langle \{ x_{ij} - u_i u_j \}_{i,j} \in \mathbb{R}^{m,n} \} \right\rangle + \sum_{i=1}^{m} u_i^2 - 1, \sum_{j=1}^{n} v_j^2 - 1 \right\rangle \]

in \( \mathbb{R}[x, u, v] = \mathbb{R}[x_1, \ldots, x_{mn}, u_1, \ldots, u_m, v_1, \ldots, v_n] \). Some tedious computations reveal the reduced Gröbner basis \( G \) of the ideal \( J \) with respect to the grevlex (and
with the corresponding \( \theta \) ordering,

\[
G = \left\{ g_1^{ij} = x_{ij} - u_i v_j : i \in [m], j \in [n] \right\} \cup \left\{ g_2 = \sum_{i=1}^{m} u_i^2 - 1, g_3 = \sum_{j=1}^{n} v_j^2 - 1 \right\}
\]

\[
\cup \left\{ g_4^{i,j,k} = x_{ij} u_k - x_{kj} u_i : 1 \leq i < k \leq m, j \in [n] \right\}
\]

\[
\cup \left\{ g_5^{i,j,k} = x_{ij} v_k - x_{ik} v_j : i \in [m], 1 \leq j < k \leq n \right\}
\]

\[
\cup \left\{ g_6^i = \sum_{j=1}^{n} x_{ij} v_j - u_i : i \in [m] \right\} \cup \left\{ g_7^j = \sum_{i=1}^{m} x_{ij} u_i - v_j : j \in [n] \right\}
\]

\[
\cup \left\{ g_8^{i,j} = \sum_{k=1}^{n} x_{ik} x_{jk} - u_i u_j : 1 \leq i < j \leq m \right\}
\]

\[
\cup \left\{ g_9^{j,k} = \sum_{k=1}^{n} x_k i u_k - v_i v_j : 1 \leq i < j \leq n \right\}
\]

\[
\cup \left\{ g_{10} = \sum_{j=1}^{n} x_{ij}^2 : 2 \leq i \leq m \right\} \cup \left\{ g_{11}^{j} = \sum_{i=1}^{m} x_{ij}^2 - v_j^2 : 2 \leq j \leq n \right\}
\]

\[
\cup \left\{ g_{12}^{i,j,k,l} = x_{ij} x_{kl} - x_{il} x_{kj} : 1 \leq i < k \leq m, 1 \leq j < l \leq n \right\}
\]

\[
\cup \left\{ g_{13} = x_{11}^2 - \sum_{i=2}^{m} \sum_{j=2}^{n} x_{ij}^2 + \sum_{i=2}^{m} u_i^2 + \sum_{j=2}^{n} v_j^2 - 1 \right\}. \quad (12)
\]

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 \( G_{M_{mn}} \) defined in \( \text{[8]} \). Notice that \( \sum_{i=1}^{m} \sum_{j=1}^{n} x_{ij}^2 - 1 = g_{13} + \sum_{i=2}^{m} g_{12}^{i} + \sum_{j=2}^{n} g_{11}^{j} \) together with \( \{g_{12}^{i,j,k,l}\} \) form the basis \( G_{M_{mn}} \).

3.1. The \( \theta_1 \)-norm in \( \mathbb{R}^{2 \times 2} \). For the sake of illustration, we consider the specific example of 2 \( \times 2 \) matrices and provide the corresponding semidefinite program for the computation of the \( \theta_1 \)-norm explicitly. Let us denote the corresponding polynomial ideal in \( \mathbb{R}[x] = \mathbb{R}[x_{11}, x_{12}, x_{21}, x_{22}] \) simply by

\[
J = J_{22} = \langle x_{12} x_{21} - x_{11} x_{22}, x_{11}^2 + x_{12}^2 + x_{21}^2 + x_{22}^2 - 1 \rangle \quad (13)
\]

The associated algebraic variety is of the form

\[
\nu_2(J) = \{ x : x_{12} x_{21} = x_{11} x_{22}, x_{11}^2 + x_{12}^2 + x_{21}^2 + x_{22}^2 - 1 \}
\]

and corresponds to the set of rank-one matrices with \( \|X\|_F = 1 \). Its convex hull consists of matrices \( X \in \mathbb{R}^{2 \times 2} \) with \( \|X\|_* \leq 1 \). According to Lemma 3.2 the Gröbner basis \( G \) of \( J \) with respect to the grevlex order is

\[
G = \{ g_1 = x_{12} x_{21} - x_{11} x_{22}, g_2 = x_{11}^2 + x_{12}^2 + x_{21}^2 + x_{22}^2 - 1 \}
\]

with the corresponding \( \theta \)-basis \( B \) of \( \mathbb{R}[x] / J \) restricted to the degree two given as

\[
B_1 = \{ 1 + J, x_{11} + J, x_{12} + J, x_{21} + J, x_{22} + J \}
\]

\[
B_2 = B_1 \cup \{ x_{11} x_{12} + J, x_{11} x_{21} + J, x_{11} x_{22} + J, x_{12}^2 + J, x_{12} x_{22} + J, x_{21}^2 + J, x_{21} x_{22} + J, x_{22}^2 + J \}.
\]
Remark 3. In compressive sensing, reconstruction of sparse signals via $\ell_1$-norm minimization is well-understood, see for example [10, 19, 22]. It is possible to provide

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

The set $B_2$ 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 $G$. For example, $x_{11}x_{12} + J$ is an element of the theta basis $B$, but $x_{11}^2 + J$ is not since $x_{11}^2$ is divisible by $LT(g_2)$.

Linearizing the elements of $B_2$ results in Table 1, where the monomials $f$ in the first row stand for an element $f + J \in B_2$. Therefore, $[x]_{B_2} = (1, x_{11}, x_{12}, x_{21}, x_{22})^T$ and the following combinatorial moment matrix $M_{B_1} (x, y)$, see Definition 2.4, is given as

$$M_{B_1} (x, y) = \begin{bmatrix} y_0 & x_{11} & x_{12} & x_{21} & x_{22} \\ x_{11} & -y_4 - y_6 - y_8 + y_0 & y_1 & y_2 & y_3 \\ x_{12} & y_1 & y_4 & y_5 \\ x_{21} & y_2 & y_3 & y_6 & y_7 \\ x_{22} & y_3 & y_5 & y_7 & y_8 \end{bmatrix}.$$  

For instance, the entry $(2, 2)$ of $[x]_{B_2}$, $[x]_{B_2}$ 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 2.3 and the fact that $g_2 \in J$. Replacing $x_{12}^2 + J$ by $y_4$, etc. as in Table 1 yields the stated expression for $M_{B_1} (x, y)_{2, 2}$.

By Theorem 2.5, the first theta body $TH_1 (J)$ is the closure of

$$Q_{B_1} (J) = \{ (x, y) \in \mathbb{R}^{B_2} : M_{B_1} (x, y) \succeq 0, y_0 = 1 \},$$

where $\pi_x$ represents the projection onto the variables $x$, i.e., the projection onto $x_{11}$, $x_{12}$, $x_{21}$, $x_{22}$. Furthermore, $\ell_1$-norm of a matrix $X \in \mathbb{R}^{2 \times 2}$ induced by the $TH_1 (J)$ and denoted as $\|\|_\theta_1$ can be computed as

$$\|X\|_{\theta_1} = \inf_t \text{ s.t. } X \in tQ_{B_1} (J)$$

which is equivalent to

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

Notice that $\text{trace}(M) = 2t$. By Theorem 3.5, the above program is equivalent to the standard semidefinite program for computing the nuclear norm of a given matrix $X \in \mathbb{R}^{m \times n}$

$$\min_{W, Z} \frac{1}{2} (\text{trace} (W) + \text{trace} (Z)) \text{ s.t. } \begin{bmatrix} W_{11} & W_{12} & X_{11} & X_{12} \\ W_{12} & W_{22} & X_{21} & X_{22} \\ X_{11} & X_{21} & Z_{11} & Z_{12} \\ X_{22} & Z_{22} & Z_{12} & Z_{22} \end{bmatrix} \succeq 0.$$
hierarchical relaxations via theta bodies of the unit $\ell_1$-norm ball. However, as in the matrix scenario discussed above, all these relaxations coincide with the unit $\ell_1$-norm ball. [57]

4. The tensor $\theta_k$-norm

Let us now turn to the tensor case and study the hierarchical closed convex relaxations of the unit tensor nuclear norm ball defined via theta bodies. Since in the matrix case all $\theta_k$-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 $\theta_1$-norm whose unit norm ball is the largest in a hierarchical sequence of relaxations. Unlike in the matrix case, the $\theta_1$-norm defines a new tensor norm, that up to the best of our knowledge has not 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 $\theta_k$-norm to a certain tensor format. We refer to e.g. [32] [52] for some background on various tensor formats. However, as we will show later, the corresponding reduced Gröbner basis with respect to the grevlex order does not depend on the choice of the tensor format. We refer to e.g. [32] [52] for some background on various tensor formats.

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

$$S_1 = \{ (\alpha, \beta) : 1 \leq \alpha_1 < \beta_1 \leq n_1, 1 \leq \beta_2 < \alpha_2 \leq n_2, 1 \leq \beta_3 \leq \alpha_3 \leq n_3 \},$$

$$S_2 = \{ (\alpha, \beta) : 1 \leq \alpha_1 \leq \beta_1 \leq n_1, 1 \leq \beta_2 < \alpha_2 \leq n_2, 1 \leq \alpha_3 \leq \beta_3 \leq n_3 \},$$

$$S_3 = \{ (\alpha, \beta) : 1 \leq \alpha_1 < \beta_1 \leq n_1, 1 \leq \alpha_2 \leq \beta_2 \leq n_2, 1 \leq \alpha_3 \leq \beta_3 \leq n_3 \},$$

$$\overline{S}_i = \{ (\alpha, \beta) : (\alpha, \beta) \in S_i \text{ and } \alpha_j \neq \beta_j, \text{ for all } j \in [3] \}, \text{ for all } i \in [3].$$

The following polynomials $f^{(\alpha, \beta)}$ in $\mathbb{R}[x] = \mathbb{R}[x_{111}, x_{112}, \ldots, x_{n_1 n_2 n_3}]$ correspond to a subset of all order-two minors of all tensor unfoldings,

$$f^{(\alpha, \beta)}(x) = x_{\alpha \lor \beta} x_{\alpha \land \beta} - x_{\alpha \land \beta}, \quad (\alpha, \beta) \in S := S_1 \cup S_2 \cup S_3$$

where $|\alpha \lor \beta| = \max \{\alpha_i, \beta_i\}$ and $|\alpha \land \beta| = \min \{\alpha_i, \beta_i\}$. In particular, the following order-two minor of $X^{(1)}$ is not contained in $\{ f^{(\alpha, \beta)} : (\alpha, \beta) \in S \}$

$$f = x_{\alpha \land \beta} - x_{\alpha \land \beta}, \quad \text{ where } \hat{\alpha} = (\alpha_1, \beta_2, \beta_3), \hat{\beta} = (\beta_1, \alpha_2, \alpha_3) \text{ and } (\alpha, \beta) \in \overline{S}_3.$$
We remark that in real algebraic geometry and commutative algebra, polynomials $f^{(\alpha, \beta)}$ are known as Hibi relations, see [33].

**Lemma 4.1.** A tensor $X \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is a rank-one, unit Frobenius norm tensor if and only if

$$g_3(X) = 0 \text{ and } f^{(\alpha, \beta)}(X) = 0 \text{ for all } (\alpha, \beta) \in S.$$  \hspace{1cm} (16)

**Proof.** Sufficiency of (16) follows directly from the definition of the rank-one unit Frobenius norm 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 $\beta_2 \leq \alpha_2$ and $\beta_3 \leq \alpha_3$, the tensors $X_{\alpha_2 \alpha_3}$ and $X_{\beta_3 \beta_3}$ satisfy

$$-X_{n_1 \alpha_2 \alpha_3} \begin{bmatrix} X_{1 \beta_3} \\ X_{2 \beta_3} \\ \vdots \\ X_{n_1 \beta_3} \end{bmatrix} + X_{n_1 \beta_3} \begin{bmatrix} X_{1 \alpha_3} \\ X_{2 \alpha_3} \\ \vdots \\ X_{n_1 \alpha_3} \end{bmatrix} = 0,$$

where we used that $f^{(\alpha, \beta)}(X) = 0$ for all $(\alpha, \beta) \in S$. From $g_3(X) = 0$ it follows that the tensor $X$ 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 $X \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is rank one if and only if all three unfoldings $X^{(1)} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, $X^{(2)} \in \mathbb{R}^{n_2 \times n_1 \times n_3}$, and $X^{(3)} \in \mathbb{R}^{n_3 \times n_1 \times n_2}$ are rank-one matrices. Notice that $f^{(\alpha, \beta)}(X) = 0$ for all $(\alpha, \beta) \in S_{\ell}$ is equivalent to the statement that the $\ell$-th unfolding $X^{(\ell)}$ is a rank-one matrix, i.e., that all its order-two minors vanish, for all $\ell \in [3]$.

In order to define relaxations of the unit tensor nuclear norm ball we introduce the polynomial ideal $J_3 \subseteq \mathbb{R}[X] = \mathbb{R}[x_{111}, x_{112}, \ldots, x_{n_1 n_2 n_3}]$ as the one generated by

$$G_3 = \left\{ f^{(\alpha, \beta)}(x) : (\alpha, \beta) \in S \right\} \cup \{ g_3(x) \},$$  \hspace{1cm} (17)

i.e., $J_3 = \langle G_3 \rangle$. Its real algebraic variety equals the set of rank-one third-order tensors with unit Frobenius norm and its convex hull coincides with the unit tensor nuclear norm ball. The next result provides the Gröbner basis of $J_3$.

**Theorem 4.2.** The basis $G_3$ defined in (17) forms the reduced Gröbner basis of the ideal $J_3 = \langle G_3 \rangle$ with respect to the grevlex order.

**Proof.** Similarly to the proof of Theorem 3.2 we need to show that $S(p, q) \rightarrow_{G_3} 0$ for all polynomials $p, q \in G_3$ whose leading terms are not relatively prime. The leading monomials with respect to the grevlex ordering are given by

$$\text{LM}(g_3) = x_{111}^2 \quad \text{and} \quad \text{LM}(f^{(\alpha, \beta)}) = x_{\alpha} x_{\beta}, \quad (\alpha, \beta) \in S.$$  

The leading terms of $g_3$ and $f^{(\alpha, \beta)}$ are always relatively prime. First we consider two distinct polynomials $f, g \in \{ f^{(\alpha, \beta)} : (\alpha, \beta) \in S_3 \}$. Let $f = f^{(\alpha, \beta)}$ and $g = f^{(\beta, \beta)}$ for $(\alpha, \beta) \in S_3$, where $S_3 = (\beta_1, \alpha_2, \beta_3)$. That is,$$

f(x) = x_{\alpha} x_{\beta} - x_{\alpha \lor \beta} x_{\alpha \land \beta}, \quad g(x) = x_{\alpha} x_{\beta} - x_{\alpha \lor \beta} x_{\alpha \land \beta}.$$


Remark 4. The above Gröbner basis \( \mathcal{G}_3 \) is obtained by taking a particular subset of all order-two minors of all three unfoldings of the tensor \( X \in \mathbb{R}^{n_1 \times n_2 \times n_3} \) (not considering the same minor twice). One might think that the \( \theta_1 \)-norm obtained in this way corresponds to a (weighted) sum of the nuclear norms of the unfoldings, which has been used in \cite{21,35} for tensor recovery. The examples of cubic tensors \( X \in \mathbb{R}^{2 \times 2 \times 2} \) presented in Table 2 show that this is not the case. Assuming that \( \theta_1 \)-norm is a linear combination of the nuclear norms of the unfoldings, there exist \( \alpha, \beta, \gamma \in \mathbb{R} \) such that \( \alpha \| X^{(1)} \|_\ast + \beta \| X^{(2)} \|_\ast + \gamma \| X^{(3)} \|_\ast = \| X \|_{\theta_1} \). From the first and the second tensor in Table 2 we obtain \( \gamma = 0 \). Similarly, the first and the third tensor, and the first and the fourth tensor give \( \beta = 0 \) and \( \alpha = 0 \), respectively. Thus, the \( \theta_1 \)-norm does not coincide with a weighted sum of the \( \theta_2 \)-norms of the unfoldings. In addition, the last tensor shows that the \( \theta_1 \)-norm does not equal maximum of the \( \theta_2 \)-norms of the unfoldings.

Theorem 3.2 states that \( \mathcal{G}_3 \) is the reduced Gröbner basis of the ideal \( J_3 \) generated by all order-two minors of all matricizations of an order-three tensor. That is, \( J_3 \) is

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|}
\hline
\( X \in \mathbb{R}^{2 \times 2 \times 2} \) & \( \| X^{(1)} \|_\ast \) & \( \| X^{(2)} \|_\ast \) & \( \| X^{(3)} \|_\ast \) & \( \| X \|_{\theta_1} \) \\
\hline
1 & \[
\begin{bmatrix}
1 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 1
\end{bmatrix}
\] & 2 & 2 & 2 \\
2 & \[
\begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 0
\end{bmatrix}
\] & 2 & 2 & \( \sqrt{2} \) \\
3 & \[
\begin{bmatrix}
1 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 1
\end{bmatrix}
\] & 2 & \( \sqrt{2} \) & 2 \\
4 & \[
\begin{bmatrix}
1 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 1
\end{bmatrix}
\] & \( \sqrt{2} \) & 2 & 2 \\
5 & \[
\begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 0
\end{bmatrix}
\] & \( \sqrt{2}+1 \) & \( \sqrt{2}+1 \) & 3 \\
\hline
\end{tabular}
\end{table}
generated by the following polynomials

\[ f_{(\alpha, \beta)}^{(1)}(x) = -x_{a_1 a_2 a_3} x_{\beta_1 \beta_2 \beta_3} + x_{a_1 \beta_2 \beta_3} x_{\beta_1 a_2 a_3}, \quad \text{for } (\alpha, \beta) \in \mathcal{T}^{(1)} \]

\[ f_{(\alpha, \beta)}^{(2)}(x) = -x_{a_1 a_2 a_3} x_{\beta_1 \beta_2 \beta_3} + x_{\beta_1 a_2 \beta_3} x_{a_1 \beta_2 a_3}, \quad \text{for } (\alpha, \beta) \in \mathcal{T}^{(2)} \]

\[ f_{(\alpha, \beta)}^{(3)}(x) = -x_{a_1 a_2 a_3} x_{\beta_1 \beta_2 \beta_3} + x_{\beta_1 \beta_2 a_3} x_{a_1 a_2 \beta_3}, \quad \text{for } (\alpha, \beta) \in \mathcal{T}^{(3)} \]

where \( \{ f_{(\alpha, \beta)}^{(k)}(x) : (\alpha, \beta) \in \mathcal{T}^{(k)} \} \) is the set of all order-two minors of the \( k \)th unfolding and

\[ \mathcal{T}^{(k)} = \{ (\alpha, \beta) : a_k \neq \beta_k, \overline{\alpha} \neq \overline{\beta}, \text{where } \overline{a}_k = \overline{\beta}_k = 0, \overline{\alpha}_\ell = \alpha_\ell, \overline{\beta}_\ell = \beta_\ell \} \]

For \((\alpha, \beta), x_\alpha^{(k)}x_\beta^{(k)} \) denotes a monomial where \( \alpha_k = a_k, \beta_k = \beta_k, \) and \( \alpha_\ell = \beta_\ell, \beta_\ell = \alpha_\ell, \) for all \( \ell \in [d] \setminus \{k\} \). Notice that \( f_{(\alpha, \beta)}^{(k)}(x) = f_{(\beta, \alpha)}^{(k)}(x) = f_{(\alpha, \beta)}^{(k)}(x) \), for all \( (\alpha, \beta) \in \mathcal{T}^{(k)} \), and all \( k \in [3] \). Let us now consider a TT-format and a corresponding notion of tensor rank. Recall that a TT-rank of an order three tensor is a vector \( r = (r_1, r_2) \) where \( r_1 = \text{rank}(X^{(1)}) \) and \( r_2 = \text{rank}(X^{(1,2)}) \). Consequently, we consider an ideal \( J_{3,TT} \) generated by all order-two minors of matricizations \( X^{(1)} \) and \( X^{(1,2)} \) of the order-3 tensor. That is, the ideal \( J_{3,TT} \) is generated by the polynomials

\[ f_{(\alpha, \beta)}^{(1)}(x) = -x_{a_1 a_2 a_3} x_{\beta_1 \beta_2 \beta_3} + x_{a_1 \beta_2 \beta_3} x_{\beta_1 a_2 a_3}, \quad \text{for } (\alpha, \beta) \in \mathcal{T}^{(1)} \]

\[ f_{(\alpha, \beta)}^{(2)}(x) = -x_{a_1 a_2 a_3} x_{\beta_1 \beta_2 \beta_3} + x_{\beta_1 a_2 \beta_3} x_{a_1 \beta_2 a_3}, \quad \text{for } (\alpha, \beta) \in \mathcal{T}^{(2)} \]

\[ f_{(\alpha, \beta)}^{(3)}(x) = -x_{a_1 a_2 a_3} x_{\beta_1 \beta_2 \beta_3} + x_{\beta_1 \beta_2 a_3} x_{a_1 a_2 \beta_3}, \quad \text{for } (\alpha, \beta) \in \mathcal{T}^{(3)} \]

where \( \mathcal{T}^{(1,2)} = \{ (\alpha, \beta) : (a_1, a_2, 0) \neq (\beta_1, \beta_2, 0), a_3 \neq \beta_3 \} \).

**Theorem 4.3.** The polynomial ideals \( J_3 \) and \( J_{3,TT} \) are equal.

**Remark 5.** As a consequence, \( G_3 \) is also the reduced Gröbner basis for the ideal \( J_{3,TT} \) with respect to the grevlex ordering.

**Proof.** Notice that \( (X^{(3)})^T = X^{(1,2)} \) and therefore

\[ \{ f_{(\alpha, \beta)}^{(3)}(x) : (\alpha, \beta) \in \mathcal{T}^{(3)} \} = \{ f_{(\alpha, \beta)}^{(1,2)}(x) : (\alpha, \beta) \in \mathcal{T}^{(1,2)} \} \]

Hence, it is enough to show that \( f_{(\alpha, \beta)}^{(2)} \in J_{3,TT} \), for all \( (\alpha, \beta) \in \mathcal{T}^{(2)} \). By definition of \( \mathcal{T}^{(2)} \), we have that \( \alpha_2 \neq \beta_2 \) and \((a_1, 0, a_3) \neq (\beta_1, 0, \beta_3)\). We can assume that \( \alpha_3 \neq \beta_3 \), since otherwise \( f_{(\alpha, \beta)}^{(2)} = f_{(\alpha, \beta)}^{(1)} \). Analogously, \( \alpha_1 \neq \beta_1 \) since otherwise \( f_{(\alpha, \beta)}^{(2)} = f_{(\alpha, \beta)}^{(1,2)} \). Consider the following polynomials

\[ f(x) = -x_{a_1 a_2 a_3} x_{\beta_1 \beta_2 \beta_3} + x_{\beta_1 a_2 \beta_3} x_{a_1 \beta_2 a_3}, \quad (\alpha, \beta) \in \mathcal{T}^{(2)} \]

\[ g(x) = -x_{\beta_1 \beta_2 a_3} x_{a_1 a_2 \beta_3} + x_{\beta_1 \beta_2 \beta_3} x_{a_1 a_2 a_3}, \quad (\beta_1, \beta_2, 0, a_1, a_2, \beta_3) \in \mathcal{T}^{(1)} \]

\[ h(x) = -x_{a_1 a_2 a_3} x_{\beta_1 \beta_2 \beta_3} + x_{a_1 a_2 \beta_3} x_{\beta_1 \beta_2 a_3}, \quad (\alpha, \beta) \in \mathcal{T}^{(1,2)} \]

Thus, we have that \( f(x) = g(x) + h(x) \in J_{3,TT} \).

### 4.2. The theta norm for general \( d \)th-order tensors

Let us now consider \( d \)th-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 \( X \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d} \) 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 \( X \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d} \) 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 to before, the entry \((\alpha_1, \alpha_2, \ldots, \alpha_d)\) of a tensor \(X \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}\) corresponds to the variable \(x_{\alpha_1, \alpha_2, \ldots, \alpha_d}\) or simply \(x_\alpha\). We aim at introducing a set of polynomials of the form

\[
f_d^{(\alpha, \beta)}(x) := -x_\alpha \wedge \beta x_\alpha \vee \beta + x_\alpha x_\beta
\]

which will generate the desired polynomial ideal. These polynomials correspond to a subset of all order-two minors of the possible \(d\)-th-order tensor matricizations. The set \(\mathcal{S}\) denotes the indices where \(\alpha\) and \(\beta\) differ. Since for an order-two minor of a matricization \(X^M\) the sets \(\alpha\) and \(\beta\) need to differ in at least two indices, \(\mathcal{S}\) is contained in

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

Given the set \(\mathcal{S}\) of different indices, we require all non-empty subsets \(\mathcal{M} \subset \mathcal{S}\) of possible indices which are “switched” between \(\alpha\) and \(\beta\) for forming the minors in \(\mathcal{S}_{[d]}\). This implies that, without loss of generality,

\[
\alpha_j > \beta_j, \quad \text{for all } j \in \mathcal{M} \\
\alpha_k < \beta_k, \quad \text{for all } k \in \mathcal{S}\setminus \mathcal{M}.
\]

That is, the same minor is obtained if we require that \(\alpha_j < \beta_j\) for all \(j \in \mathcal{M}\) and \(\alpha_k > \beta_k\) for all \(k \in \mathcal{S}\setminus \mathcal{M}\) since the set of all two-minors of \(X^M\) coincides with the set of all two-minors of \(X^{\mathcal{S}\setminus \mathcal{M}}\).

For \(S \in \mathcal{S}_{[d]}\), we define \(e_S := \min\{p : p \in S\}\). The set \(\mathcal{M}\) corresponds to an associated matricization \(X^M\). The set of possible subsets \(\mathcal{M}\) is given as

\[
\mathcal{P}_S = \begin{cases} 
\{ \mathcal{M} \subset S : |\mathcal{M}| \leq \lfloor \frac{|S|}{2} \rfloor \} \setminus \{\emptyset\}, & \text{if } |S| \text{ is odd,} \\
\{ \mathcal{M} \subset S : |\mathcal{M}| \leq \lfloor \frac{|S| - 1}{2} \rfloor \} \cup \{ \mathcal{M} \subset S : |\mathcal{M}| = \lfloor \frac{|S|}{2} \rfloor, e_S \in \mathcal{M} \} \setminus \{\emptyset\}, & \text{otherwise.}
\end{cases}
\]

Notice that \(\mathcal{P}_S \cup \mathcal{P}_{S^c} \cup \{\emptyset\} \cup \mathcal{S}\) with \(\mathcal{P}_{S^c} := \{\mathcal{M} : S^c \setminus \mathcal{M} \in \mathcal{P}_S\}\) forms the power set of \(\mathcal{S}\). The constraint on the size of \(\mathcal{M}\) in the definition of \(\mathcal{P}_S\) is motivated by the fact that the role of \(\alpha\) and \(\beta\) can be switched and lead to the same polynomial \(f_d^{(\alpha, \beta)}\).

Thus, for \(S \in \mathcal{S}_{[d]}\) and \(\mathcal{M} \in \mathcal{P}_S\), we define a set

\[
\mathcal{T}^{S, M}_d := \{(\alpha, \beta) : \alpha_i = \beta_i, \text{ for all } i \notin S \\
\alpha_j > \beta_j, \text{ for all } j \in \mathcal{M} \\
\alpha_k < \beta_k, \text{ for all } k \in \mathcal{S}\setminus \mathcal{M}\}.
\]

For notational purposes, we define

\[
\{f_d^S\} = \bigcup_{\mathcal{M} \in \mathcal{P}_S} \{f_d^{(\alpha, \beta)} : (\alpha, \beta) \in \mathcal{T}^{S, M}_d\} \text{ for } S \in \mathcal{S}_{[d]}.
\]

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

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

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

\[
G_d = \bigcup_{S \in \mathcal{S}_{[d]}} \{f_d^S\} \cup \{g_d\} \subset \mathbb{R}[x] = \mathbb{R}[x_{111, 112, \ldots, x_{n_1 n_2 \ldots n_d}}],
\]

i.e., \(J_d = (G_d)\). As in the special case of the third-order tensors, not all second order minors corresponding to all matricizations are contained in the generating set \(G_d\) due to the condition \(i_k < j_k\) for all \(k \in \mathcal{S}\) in the definition of \(G_d^S\). Nevertheless all second order minors are contained in the ideal \(J_d\) as will also be revealed by
the proof of Theorem 4.4 below. For instance, \( h(x) = -x_{1234}x_{2343} + x_{1243}x_{2334} \) corresponding to a minor of the matricization \( X^M \) for \( M = \{1, 2\} \) – does not belong to \( \mathcal{G}_4 \), but it does belong to the ideal \( \mathcal{J}_4 \). Moreover, it is straightforward to verify that all polynomials in \( \mathcal{G}_d \) differ from each other.

The algebraic variety of \( \mathcal{J}_d \) consists of all rank-one unit Frobenius norm order-\( d \) tensors as desired, and its convex hull yields the tensor nuclear norm ball.

**Theorem 4.4.** The set \( \mathcal{G}_d \) forms the reduced Gröbner basis of the ideal \( \mathcal{J}_d \) with respect to the grevlex order.

**Proof.** Again, we use Buchberger’s criterion stated in Theorem A.7. First notice that the polynomials \( g_d \) and \( f_d(\alpha, \beta) \) are always relatively prime, since \( \text{LM}(g_d) = x_1^d \ldots 1 \) and \( \text{LM}(f_d(\alpha, \beta)) = x_\alpha x_\beta \) for \( (\alpha, \beta) \in T_d^{M,S} \), where \( S \in S_d \) and \( M \in \mathcal{P}_S \). Therefore, we need to show that \( S(f_1, f_2) \rightarrow \mathcal{G}_d \) 0, for all \( f_1, f_2 \in \mathcal{G}_d \setminus \{g_d\} \) 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 \( \text{LM}(f_1) \neq \text{LM}(f_2) \). If these leading monomials are not relatively prime, the \( S \)-polynomial is of the form

\[
S(f_1, f_2) = x_{\alpha_1} x_{\alpha_2} x_{\alpha_3} - x_{\alpha_1} x_{\alpha_2} x_{\alpha_3}^3
\]

with \( \{\alpha_1, \alpha_2, \alpha_3\} = \{\beta_1^k, \beta_2^k, \beta_3^k\} \) for all \( k \in [d] \).

The step-by-step procedure of the division algorithm for our scenario is presented in Algorithm 2. We will show that the algorithm eventually stops and that step 2) is feasible, i.e., that there always exist \( k \) and \( \ell \) such that line 7 of Algorithm 2 holds – provided that \( S^i \neq 0 \). (In fact, the purpose of the algorithm is to achieve the condition that in the \( i \)th iteration of the algorithm \( \beta_1^{k_i} \leq \beta_2^{k_i} \leq \beta_3^{k_i} \) for all \( k \in [d] \). This will show then that \( S(f_1, f_2) \rightarrow \mathcal{G}_d \) 0.

**Algorithm 2** The division algorithm on the ideal \( \langle \mathcal{G}_d \rangle \).

**Input:** polynomials \( f_1, f_2 \in \mathcal{G}_d \)

\[ S^0 = S(f_1, f_2) = x_{\alpha_1} x_{\alpha_2} x_{\alpha_3} - x_{\alpha_1} x_{\alpha_2} x_{\alpha_3}^3, \quad i = 0 \]

while \( S^i \neq 0 \) do

1) Let \( \text{LM}(S^i) = x_{\alpha_1} x_{\alpha_2} x_{\alpha_3}^2 \) and \( \text{LNM}(S^i) = \text{LM}(S^i) = |S^i| - \text{LT}(S^i)| \)

2) Find indices \( \alpha^{1,i}, \alpha^{2,i}, \alpha^{3,i} \in \{\hat{\alpha}^{1,i}, \hat{\alpha}^{2,i}, \hat{\alpha}^{3,i}\} \) such that there exist at least one \( k \) and at least one \( \ell \) for which

\[
\alpha_1^{k_i} \leq \alpha_2^{k_i} \quad \text{and} \quad \alpha_1^{\ell_i} > \alpha_2^{\ell_i} \quad \text{s.t.} \quad M_i := \{\ell \in [d] : \alpha_1^{\ell_i} > \alpha_2^{\ell_i}\} \in \mathcal{P}_S,
\]

where \( S := \{k \in [d] : \alpha_1^{k_i} \neq \alpha_2^{k_i}\} \) and let \( \alpha^{3,i} \) be the remaining index in \( \{\hat{\alpha}^{1,i}, \hat{\alpha}^{2,i}, \hat{\alpha}^{3,i}\} \setminus \{\alpha^{1,i}, \alpha^{2,i}\} \).

3) Divide \( S^i \) by \( f_d(\alpha^{1,i}, \alpha^{2,i}) = x_{\alpha_1^{1,i}} x_{\alpha_2^{2,i}} - x_{\alpha_1^{1,i} \land \alpha_2^{2,i} \lor \alpha_3^{3,i}} \) to obtain

\[
S^i = \text{LC}(S^i) \left[ x_{\alpha_2^{3,i}} (-x_{\alpha_1^{1,i} \land \alpha_2^{2,i} \lor \alpha_3^{3,i}} + x_{\alpha_1^{1,i} \land \alpha_2^{2,i}}) + x_{\alpha_1^{1,i} \land \alpha_2^{2,i} \lor \alpha_3^{3,i} - \text{LNM}(S^i)} \right].
\]

4) Define

\[
S^{i+1} := x_{\alpha_1^{1,i} \land \alpha_2^{2,i} \lor \alpha_3^{3,i}} - \text{LNM}(S^i).
\]

5) \( i := i + 1 \)

end while

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 \( f_1(x) := f_{4(121,2212)}(x) = -x_{1112}x_{2223} + x_{1212}x_{2113} \in G_4 \) (with the corresponding sets \( \mathcal{S} = \{1, 2, 3, 4\}, \mathcal{M} = \{2\} \) and \( f_2(x) := f_{4(3311,2123)}(x) = -x_{1112}x_{3323} + x_{3311}x_{2123} \in G_4 \) (with the corresponding sets \( \mathcal{S} = \{1, 2, 3, 4\}, \mathcal{M} = \{1, 2\} \)). We will show that \( S(f_1, f_2) = -x_{1112}x_{2223}x_{3311} + x_{1212}x_{2111}x_{3323} \to g_4, 0 \) 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 \( \text{LM}(S^0) = x_{1111}x_{2223}x_{3311} \), the leading coefficient is \( \text{LC}(S^0) = -1 \), and the non-leading monomial is \( \text{NLM}(S^0) = x_{1212}x_{2111}x_{3323}. \)

Among the two options for choosing a pair of indexes \((\alpha^{1,0}, \alpha^{2,0})\) in step 2, we decide to take \( \alpha^{1,0} = 1112 \) and \( \alpha^{2,0} = 3311 \) which leads to the set \( \mathcal{M}_0 = \{4\}. \)

The polynomial \( x_{\alpha^{1,0}}x_{\alpha^{2,0}} - x_{\alpha^{1,0}}x_{\alpha^{2,0}}x_{\alpha^{1,0}}x_{\alpha^{2,0}} \) then equals the polynomial \( f_{4(1112,2331)}(x) = -x_{1112}x_{3312} + x_{1112}x_{3311} \in G_4 \) and we can write

\[
S^0 = -1 \cdot \left( x_{3323}(-x_{1111}x_{3312} + x_{1111}x_{3311}) + x_{1111}x_{2223}x_{3312} - x_{1212}x_{2111}x_{3323} \right).
\]

The leading and non-leading monomials of \( S^1 \) are \( \text{LM}(S^1) = x_{1111}x_{2223}x_{3312} \) and \( \text{NLM}(S^1) = x_{1212}x_{2111}x_{3323} \), respectively, while \( \text{LC}(S^1) = 1 \). The only option for a pair of indices as in line 7 of Algorithm 2 is \( \alpha^{1,1} = 3312, \alpha^{2,1} = 2223 \), so that the set \( \mathcal{M}_1 = \{1, 2\} \).

The divisor \( x_{\alpha^{1,1}}x_{\alpha^{2,1}} - x_{\alpha^{1,1}}x_{\alpha^{2,1}}x_{\alpha^{1,1}}x_{\alpha^{2,1}} \) equals \( f_{4(3323,2220)}(x) = -x_{1212}x_{3323} + x_{3312}x_{2223} \in G_4 \) and we obtain

\[
S^1 = 1 \cdot \left( x_{1111}(-x_{1212}x_{3323} + x_{2223}x_{3312}) + x_{1111}x_{2223}x_{3312} - x_{1212}x_{2111}x_{3323} \right) = S^2 \]

The index sets of the monomial \( x_{\alpha^{1,1}}x_{\alpha^{2,1}}x_{\alpha^{3}} = x_{1111}x_{2222}x_{3323} \) in \( S^2 \) satisfy

\[
\alpha_1^{1,1} \leq \alpha_2^{2,1} \leq \alpha_3 \quad \text{for all } k \in [4]
\]

and therefore it is the non-leading monomial of \( S^2 \), i.e., \( \text{NLM}(S^2) = x_{1111}x_{2211}x_{3323}. \)

Thus, \( \text{LM}(S^2) = x_{1212}x_{2111}x_{3323} \) and \( \text{LC}(S^2(f_1, f_2)) = -1 \). Now the only option for a pair of indices as in step 2 is \( \alpha^{1,2} = 2111 \), \( \alpha^{2,2} = 1212 \) with \( \mathcal{M}_2 = \{1\} \). This yields

\[
S^2 = -1 \cdot \left( x_{3323}(-x_{1111}x_{2212} + x_{2111}x_{1212}) + x_{1111}x_{2223}x_{3323} - x_{1111}x_{2212}x_{3323} \right) = S^3 = 0.
\]

Thus, the division algorithm stops and we obtained after three steps

\[
S(f_1, f_2) = S^0 = \text{LC}(S^0)x_{2223}f_{4(1112,2331)}(x) + \text{LC}(S^0) \text{LC}(S^1)x_{1111}f_{4(3311,2223)}(x) + \text{LC}(S^0) \text{LC}(S^1) \text{LC}(S^2)x_{3323}f_{4(2111,1212)}(x).
\]

Thus, \( S(f_1, f_2) \to g_4, 0. \)

Let us now return to the general proof. We first show that there always exist indexes \( \alpha_1, \alpha_2, \alpha_i \) satisfying line 7 of Algorithm 2 unless \( S^i = 0 \). We start by setting \( x^h = x_{\alpha_1}x_{\alpha_2}x_{\alpha_i} \) and with \( x_{\alpha_1} \geq x_{\alpha_2} \geq x_{\alpha_i} \) to be the leading monomial and \( x_{\alpha_1} \) to be the non-leading monomial of \( S^i \). The existence of a polynomial \( h \in G_d \) such that \( \text{LM}(h) \) divides \( \text{LM}(S^i) = x_{\alpha_1}x_{\alpha_2}x_{\alpha_i} \) is equivalent to the existence of \( \alpha_1, \alpha_2, \alpha_i \in \{1, 2, \ldots, d\} \) such that there exists at least one \( k \) and at least one \( l \) for which \( \alpha_k^1 < \alpha_k^2 < \alpha_k^i \) and \( \alpha_k^i > \alpha_l^j \). If such pair does not exist in iteration \( i \), we have

\[
\alpha_k^1 \leq \alpha_k^2 \leq \alpha_k^3 \quad \text{for all } k \in [d].
\]
We claim that this cannot happen if $S^i \neq 0$. In fact, (19) would imply that the monomial $x^{\alpha_i} = x_1^{\alpha_1,i} \cdots x_d^{\alpha_d,i}$ is the smallest monomial $x_1 x_2 \cdots x_d$ (with respect to the grevlex order) which satisfies
\[
\{ \alpha_k, \gamma_k, \eta_k \} = \{ \alpha_k^{1,i}, \alpha_k^{2,i}, \alpha_k^{3,i} \} \quad \text{for all } k \in [d].
\]
However, then $x^{\alpha_i}$ would not be the leading monomial by definition of the grevlex order, which leads to a contradiction. Hence, we can always find indices $\alpha^{1,i}, \alpha^{2,i}, \alpha^{3,i}$ satisfying line 7 in step 2) of Algorithm 2 unless $S^0 = 0$.

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 $\alpha^{1,0}, \alpha^{2,0}, \alpha^{3,0}$ as in step 2) of Algorithm 2. Then we divide the polynomial $S^0$ by a polynomial $h \in \mathcal{G}_d$ such that $\text{LM}(h) = x_1^{\alpha_1,0} x_2^{\alpha_2,0}$. The polynomial $h \in \mathcal{G}_d$ is defined as in step 3) of the algorithm, i.e.,
\[
h(x) = f_d^{(\alpha^{1,0}, \alpha^{2,0})} = x_1^{\alpha_1,0} x_2^{\alpha_2,0} - x_1^{\alpha_1,0} x_2^{\alpha_2,0} \in \mathcal{G}_d.
\]

The division of $S^0$ by $h$ results in
\[
S^0 = \text{LC}(S^0) \left( x_1^{\alpha_1,0} \cdot f_d^{(\alpha^{1,0}, \alpha^{2,0})} + x_1^{\alpha_1,0} x_2^{\alpha_2,0} - \text{NLM}(S^0) \right).
\]

Note that by construction
\[
[\alpha^{1,0} \wedge \alpha^{2,0}]_k \leq [\alpha^{1,0} \vee \alpha^{2,0}]_k \quad \text{for all } k \in [d]. \tag{20}
\]

If $S^1 \neq 0$, then in the following iteration $i = 1$ we can assume $\text{LM}(S^1) = x_1^{\alpha_1,1} x_2^{\alpha_2,1} x_3^{\alpha_3,0}$. Due to (20), a pair $\alpha^{1,1}, \alpha^{2,1}$ as in line 7 of Algorithm 2 can be either $\alpha^{1,0} \wedge \alpha^{2,0}, \alpha^{3,0}$ or $\alpha^{1,0} \vee \alpha^{2,0}, \alpha^{3,0}$. Let us assume the former. Then this iteration results in
\[
S^1 = \text{LC}(S^1) \left( x_1^{\alpha_1,1} \cdot f_d^{(\alpha^{1,1}, \alpha^{2,1})} + x_1^{\alpha_1,1} x_2^{\alpha_2,1} x_3^{\alpha_3,0} - \text{NLM}(S^0) \right) = S^2
\]
with
\[
[\alpha^{3,1} \wedge \alpha^{2,1}]_k \leq [\alpha^{3,1}]_k \cup [\alpha^{1,1} \vee \alpha^{2,1}]_k \quad \text{for all } k \in [d], \quad \text{and } x_3^{3,1} = x_3^{\alpha_3,0}.
\]

Next, if $S^2 \neq 0$ and $\text{LM}(S^2) = x_1^{\alpha_1,2} x_2^{\alpha_2,2} x_3^{\alpha_3,1}$ then a pair of indices satisfying line 7 of Algorithm 2 must be $\alpha^{1,1} \vee \alpha^{2,1}, \alpha^{3,1}$ so that the iteration ends up with
\[
S^2 = \text{LC}(S^2) \left( x_1^{\alpha_1,2} \cdot f_d^{(\alpha^{1,2}, \alpha^{2,2})} + x_1^{\alpha_1,2} x_2^{\alpha_2,2} x_3^{\alpha_3,1} - \text{NLM}(S^0) \right) = S^3
\]
such that
\[
[\alpha^{3,2}]_k \leq [\alpha^{1,2} \wedge \alpha^{2,2}]_k \leq [\alpha^{1,2} \vee \alpha^{2,2}]_k \quad \text{for all } k \in [d], \quad \text{and } x_3^{3,2} = x_3^{\alpha_3,1}.
\]
Thus, in iteration $i = 3$ the leading monomial $\text{LM}(S^3)$ must be $\text{NLM}(S^0)$ (unless $S^3 = 0$).

A similar analysis can be performed on the monomial $\text{NLM}(S^0)$ and therefore the algorithm stops after at most 6 iterations. The division algorithm results in
\[
S(f_1, f_2) = \sum_{i=0}^{p} \left( \prod_{j=0}^{i} \text{LC}(S^j) \right) x_1^{\alpha_{1,i}} \cdot f_d^{(\alpha^{1,i}, \alpha^{2,i})},
\]
where $f_d^{(\alpha^{1,i}, \alpha^{2,i})} = -x_1^{\alpha_1,1} x_2^{\alpha_2,1} x_3^{\alpha_3,0} + x_1^{\alpha_1,1} x_2^{\alpha_2,1} \in \mathcal{G}_d$ and $p \leq 5$. All the cases that we left out above are treated in a similar way. This shows that $\mathcal{G}_d$ is a Gröbner basis of $J_d$. 


In order to show that \( \mathcal{G}_d \) is the reduced Gröbner basis of \( J_d \), first notice that 
\( \text{LC}(g) = 1 \) 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 
\( f_d^{(\alpha^3, \beta^3)} \), \( g_d^{(\alpha^2, \beta^2)} \) \( \in \mathcal{G}_d \) (related to \( S_1, M_1 \) and \( S_2, M_2 \), respectively) it holds that 
\( \text{LM}(f_d^{(\alpha^3, \beta^3)}) \neq \text{LM}(g_d^{(\alpha^2, \beta^2)}) \) with \( (\alpha^k, \beta^k) \in \tau_S^{(3, \ell)} \) 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.

We define the tensor \( \theta_k \)-norm analogously to the matrix scenario.

**Definition 4.5.** The tensor \( \theta_k \)-norm, denoted by \( \| \cdot \|_{\theta_k} \), is the norm induced by the \( k \)-theta body \( \mathbf{T}_k(J_d) \), i.e.,
\[
\| \mathbf{X} \|_{\theta_k} = \inf \{ r : \mathbf{X} \in r \mathbf{T}_k(J_d) \}.
\]
The \( \theta_k \)-norm can be computed with the help of Theorem 2.5, i.e., as
\[
\| \mathbf{X} \|_{\theta_k} = \min t \quad \text{subject to} \quad \mathbf{X} \in t \mathbf{Q}_{\mathbf{B}_k}(J_d).
\]
Given the moment matrix \( \mathbf{M}_{\mathbf{B}_k}[\mathbf{y}] \) associated with \( J_d \), this minimization program is equivalent to the semidefinite program
\[
\min_{t \in \mathbb{R}} \quad t \quad \text{subject to} \quad \mathbf{M}_{\mathbf{B}_k}[\mathbf{y}] \succeq 0, y_0 = t, \mathbf{y}_{\mathbf{B}_i} = \mathbf{X}.
\]

We have focused 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 \( d \)-th order tensor \( \mathbf{X} \) is a \( d \)-dimensional vector \( \mathbf{r}_{\text{HOSVD}} = (r_1, r_2, \ldots, r_d) \) such that \( r_i = \text{rank}(\mathbf{X}(i)) \) for all \( i \in [d] \), see [28]. Thus, we can define an ideal \( J_d, \text{HOSVD} \) generated by all second order minors of unfoldings \( \mathbf{X}(k) \), for \( k \in [d] \).

The tensor train (TT) decomposition is another popular approach for tensor computations. The corresponding TT-rank of a \( d \)-th order tensor \( \mathbf{X} \) is a \((d - 1)\)-dimensional vector \( \mathbf{r}_{\text{TT}} = (r_1, r_2, \ldots, r_{d-1}) \) such that \( r_i = \text{rank}(\mathbf{X}(\ldots,i\ldots)) \), \( i \in [d - 1] \), see [14] for details. 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 \( J_d, \text{TT} \).

**Theorem 4.6.** The polynomial ideals \( J_d \), \( J_d, \text{HOSVD} \), and \( J_d, \text{TT} \) are equal, for all \( d \geq 3 \).

**Proof.** Let \( \tau \subset [d] \) represent a matricization. Similarly to the case of order-three tensors, for \( (\alpha, \beta) \in \mathbb{N}^{2d} \), \( x_{\alpha^r x_{\beta^r}} \) denotes the monomial where \( \alpha^r_k = \alpha_k \), \( \beta^r_k = \beta_k \) for all \( k \in \tau \) and \( \alpha^r_\ell = \beta_\ell = \alpha_\ell \) for all \( \ell \in \tau^c = [d] \setminus \tau \). Moreover, \( x_{\alpha^r x_{\beta^r} x_{\beta^r}} \) denotes the monomial where \( \alpha^r_0 = \alpha_k \), \( \beta^r_0 = \beta_k \) for all \( k \in \tau \) and \( \alpha^r_0 = \beta^r_0 = 0 \) for all \( \ell \in \tau^c = [d] \setminus \tau \). The corresponding order-two minors are defined as
\[
f_{\alpha, \beta}^{(r)}(\mathbf{x}) = -x_{\alpha^r x_{\beta^r}} + x_\alpha x_\beta, \quad (\alpha, \beta) \in \mathbf{T}_2.
\]
We define the set \( \mathbf{T}_2 \) as
\[
\mathbf{T}_2 = \left\{ (\alpha, \beta) : \alpha^{r, 0} \neq \beta^{r, 0}, \alpha^{r, 0} \neq \beta^{r, 0} \right\}.
\]
Similarly as in the case of order-three tensors, notice that \( f_{\alpha, \beta}^{(r)}(\mathbf{x}) = f_{\beta, \alpha}^{(r)}(\mathbf{x}) = -f_{\alpha, \beta}^{(r)}(\mathbf{x}) = -f_{\beta, \alpha}^{(r)}(\mathbf{x}) \), for all \( (\alpha, \beta) \in \mathbf{T}_2 \). First, we show that \( J_d = J_d, \text{HOSVD} \) by showing that \( f_{\alpha, \beta}^{(r)}(\mathbf{x}) \in J_d, \text{HOSVD} \), for all \( (\alpha, \beta) \in \mathbf{T}_2 \) and all \( |\tau| \geq 2 \). Without loss of generality, we can assume that \( \alpha_i \neq \beta_i \), for all \( i \in \tau \) since otherwise we can consider the matricization \( \tau \setminus \{ i : \alpha_i = \beta_i \} \). Additionally, by definition of
Algorithm 3 Algorithm for proving that \(J_d = J_{d,TT}\)

**Input:** An ideal \(J_{d,TT} \in \mathbb{R}[x]\), polynomial \(f_{(\alpha,\beta)}(x)\) with \(\alpha^0 = \alpha, \beta^0 = \beta, \tau = \{t_1, t_2, \ldots, t_k\}\), where \(k \geq 2\) for \(i = 1, \ldots, k\) do

- Define \(\alpha^i\) and \(\beta^i\) as
  \[
  \alpha^i_j := \begin{cases} 
  \beta^i_{j-1} & \text{if } j = t_i, \\
  \alpha^i_{j-1} & \text{otherwise}
  \end{cases}
  \quad \text{and} \quad
  \beta^i_j := \begin{cases} 
  \alpha^i_{j-1} & \text{if } j = t_i, \\
  \beta^i_{j-1} & \text{otherwise}
  \end{cases}
  \]

- Define polynomial \(g_i(x) := -x_{\alpha^{i-1}}x_{\beta^{i-1}} + x_{\alpha^i}x_{\beta^i}\).

end for

**Output:** Polynomials \(g_1, g_2, \ldots, g_k\).

that \(f_{(\alpha,\beta)}^{(k)} \in J_{d,TT}\), for all \((\alpha, \beta) \in \mathcal{T}^{(k)}\) and all \(k \in [d]\). By definition of \(J_{d,TT}\) this is true for \(k = 1\). Fix \(k \in \{2, 3, \ldots, d\}\), \((\alpha, \beta) \in \mathcal{T}^{(k)}\) and consider a polynomial \(f(x) = f_{(\alpha,\beta)}^{(k)}(x)\) corresponding to the second order minor of the matricization \(X^{(k)}\). By definition of \(\mathcal{T}^{(k)}\), \(\alpha_k \neq \beta_k\) and there exists an index \(i \in [d]\setminus \{k\}\) such that \(\alpha_i = \beta_i\). Assume that \(i > k\). Define the polynomials \(g(x) \in \mathcal{R}^{(1,2,\ldots,k),:} := \{f_{(\alpha,\beta)}^{(1,2,\ldots,k)}(x) : (\alpha, \beta) \in \mathcal{T}^{(1,2,\ldots,k)}\}\) and \(h(x) \in \mathcal{R}^{(1,2,\ldots,k-1),:} := \{f_{(\alpha,\beta)}^{(1,2,\ldots,k-1)}(x) : (\alpha, \beta) \in \mathcal{T}^{(1,2,\ldots,k-1)}\}\) as

\[
g(x) = -x_{\alpha}x_{\beta} + x_{\alpha(1,2,\ldots,k)}x_{\beta(1,2,\ldots,k)}
\]

\[
h(x) = -x_{\alpha(1,2,\ldots,k)}x_{\beta(1,2,\ldots,k-1)} + x_{\alpha(1,2,\ldots,k)}x_{\beta(1,2,\ldots,k-1)}
\]

Since \(x_{\alpha(1,2,\ldots,k)}(1,2,\ldots,k-1) \neq x_{\beta(1,2,\ldots,k)}(1,2,\ldots,k-1)\), we have \(f(x) = g(x) + h(x)\) and thus \(f \in J_{d,TT}\). If \(i < k\) notice that \(f(x) = g_1(x) + h_1(x)\), where

\[
g_1(x) = -x_{\alpha}x_{\beta} + x_{\alpha(1,2,\ldots,k-1)}x_{\beta(1,2,\ldots,k-1)} \in \mathcal{R}^{(1,2,\ldots,k-1),:}
\]

\[
h_1(x) = -x_{\alpha(1,2,\ldots,k-1)}x_{\beta(1,2,\ldots,k-1)} + x_{\alpha(1,2,\ldots,k-1)}x_{\beta(1,2,\ldots,k-1)}
\]

\[
= -x_{\alpha(1,2,\ldots,k)}x_{\beta(1,2,\ldots,k)} + x_{\alpha(k)}x_{\beta(k)} \in \mathcal{R}^{(1,2,\ldots,k),:}
\]

\[\Box\]

**Remark 6.** Fix a decomposition tree \(T_j\) which generates a particular HT-decomposition and consider the ideal \(J_{d,HT,T_j}\) generated by all second order minors corresponding to the matricizations induced by the tree \(T_j\). In a similar way as above, one can obtain that \(J_{d,HT,T_j}\) equals to \(J_d\).
5. CONVERGENCE OF THE UNIT-$\theta_k$-NORM BALLS

In this section we show the following result on the convergence of the unit $\theta_k$-balls.

**Theorem 5.1.** The theta body sequence of $J_d$ converges asymptotically to the $\Theta \theta_k$ balls, i.e.,

$$\bigcap_{k=1}^{\infty} \Theta \theta_k (J_d) = \text{conv} \left( \nu \mathbb{R} (J_d) \right).$$

To prove Theorem 5.1 we use the following result presented in [2] which is a consequence of Schmüdgen's Positivstellensatz.

**Theorem 5.2.** Let $J$ be an ideal such that $\nu \mathbb{R} (J)$ is compact. Then the theta body sequence of $J$ converges to the convex hull of the variety $\nu \mathbb{R} (J)$, in the sense that

$$\bigcap_{k=1}^{\infty} \Theta \theta_k (J) = \text{conv} \left( \nu \mathbb{R} (J) \right).$$

Proof of Theorem 5.1. The set $\nu \mathbb{R} (J_d)$ is the set of rank-one tensors with unit Frobenius norm which can be written as $\nu \mathbb{R} (J_d) = A_1 \cap A_2$ where

- $A_1 = \{ X \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d} : \text{rank}(X) = 1 \}$,
- $A_2 = \{ X \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d} : \|X\|_F = 1 \}$.

It is well-known that $A_1$ is closed [11, discussion before Definition 2.2] and since $A_2$ is clearly compact, $\nu \mathbb{R} (J_d)$ is compact. Therefore, the result follows from Theorem 5.2. □

6. COMPUTATIONAL COMPLEXITY

The computational complexity of the semidefinite programs for computing the $\theta_1$-norm of a tensor or for minimizing the $\theta_1$-norm subject to a linear constraint depends polynomially on the number of variables, i.e., on the size of $B_{2k}$, and on the dimension of the moment matrix $M$. We claim that the overall complexity scales polynomially in $n$, where for simplicity we consider $d$th-order tensors in $\mathbb{R}^{n_1 \times \cdots \times n}$. Therefore, in contrast to tensor nuclear norm minimization which is NP-hard for $d \geq 3$, tensor recovery via $\theta_1$-norm minimization is tractable.

Indeed, the moment matrix $M$ 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 $X \in \mathbb{R}^{n_1 \times \cdots \times n}$, then the number of the variables is at most $\frac{a(a+1)}{2} \sim O(a^2)$ which is polynomial in $a$. (A more precise counting does not give a substantially better estimate.)

7. NUMERICAL EXPERIMENTS

Let us now empirically study the performance of low rank tensor recovery via $\theta_1$-norm minimization via numerical experiments, where we concentrate on third-order tensors. Given measurements $b = \Phi (X)$ of a low rank tensor $X \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, where $\Phi : \mathbb{R}^{n_1 \times n_2 \times n_3} \rightarrow \mathbb{R}^m$ is a linear measurement map, we aim at reconstructing $X$ as the solution of the minimization program

$$\min \| Z \|_{\theta_1} \quad \text{subject to} \quad \Phi (Z) = b. \quad (22)$$

As outlined in Section 2, the $\theta_1$-norm of a tensor $Z$ can be computed as the minimizer of the semidefinite program

$$\min_{t, y} t \quad \text{subject to} \quad M (t, y, Z) \succeq 0,$$
TABLE 3. The matrices involved in the definition of the moment matrix $M(t, y, X)$. Due to the symmetry only the upper triangle part of the matrices is specified. The other non-specified entries of the matrices $M \in \mathbb{R}^{(n_1 n_2 n_3 + 1) \times (n_1 n_2 n_3 + 1)}$ from the first column are equal to zero. The matrix $M$ corresponds to the element $q + J_3$ of the $\theta$-basis specified in the second column. The index $I = (i, \hat{i}, j, \hat{j}, k, \hat{k})$ is in the range of the last column. The function $f : \mathbb{N}^3 \to \mathbb{N}$ is defined as $f(i, j, k) = (i - 1)n_2n_3 + (j - 1)n_3 + k + 1$.

| $\theta$-basis position $(p, q)$ in the matrix $M_{pq}$ | Range of $i, \hat{i}, j, \hat{j}, k, \hat{k}$ |
|--------------------------------------------------------|------------------------------------------|
| $M_{jk}$ $x_{ijk}$ $(1, 1), (2, 2)$ | 1 $i \in [n_1], j \in [n_2], k \in [n_3]$ |
| $M^*_jk$ $x_{ijk}$ $(1, f(i, j, k))$ | 1 $i \in [n_1], j \in [n_2], k \in [n_3]$ |
| $M^*_jk$ $(2, 2)$ | 1 $\{i \in [n_1], j \in [n_2], k \in [n_3]\}$ |
| $M^*_jk$ $(i, j, k)$ | 1 $\{i = j = k = 1\}$ |
| $M^*_jk$ $(i, j, k)$ | 1 $i < i, j < \hat{j}, k < \hat{k}$ |
| $M^*_jk$ $(i, j, k)$ | 1 $i < i, j < \hat{j}, k < \hat{k}$ |
| $M^*_jk$ $(i, j, k)$ | 1 $i < i, j \in [n_2], k < \hat{k}$ |
| $M^*_jk$ $(i, j, k)$ | 1 $i < i, j \in [n_2], k < \hat{k}$ |
| $M^*_jk$ $(i, j, k)$ | 1 $i < i, j \in [n_2], k < \hat{k}$ |

where $M(t, y, X) = M_{\mathcal{B}_3}(t, X, y)$ is the moment matrix of order 1 associated to the ideal $J_3$, see Theorem 4.2. This moment matrix for $J_3$ is explicitly given by

$$M(t, y, X) = tM_0 + \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} X_{ijk} M_{ijk} + \sum_{p=2}^{9} \sum_{q=1}^{[M^p]} y_t M^p_{\mathcal{B}_3(q)},$$

where $t = \sum_{i=2}^{p-1} [M^r] + q$, $M^p = \{M^p_I\}$, and the matrices $M_0, M_{ijk}$ and $M^p$ are provided in Table 3. For $p \in \{2, 3, \ldots, 9\}$, the function $h_p$ denotes an arbitrary but fixed bijection $\{1, 2, \ldots, [M^p]\} \rightarrow \{(i, \hat{i}, j, \hat{j}, k, \hat{k})\}$, where $\hat{i} = (i, i, j, j, k, k)$ is in the range of the last column of Table 3. As discussed in Section 4 for the general case, the $\ell_1$-norm minimization problem (22) is then equivalent to the semidefinite program

$$\min t \quad \text{subject to} \quad M(t, y, Z) \succeq 0 \quad \text{and} \quad \Phi(Z) = b. \quad (23)$$

For our experiments, the linear mapping is defined as $(\Phi(X))_k = \langle X, \Phi_k \rangle$, $k \in [m]$, with independent Gaussian random tensors $\Phi_k \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, i.e., all entries of $\Phi_k$ are independent $\mathcal{N}(0, \frac{1}{m})$ random variables. We choose tensors $X \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ of rank one as $X = u \otimes v \otimes w$, where each entry of the vectors $u$, $v$, and $w$ is taken independently from the normal distribution $\mathcal{N}(0, 1)$. Tensors $X \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ of rank two are generated as the sum of two random rank-one
With $\Phi$ and $X$ given, we compute $b = \Phi(X)$, run the semidefinite program (23) and compare its minimizer with the original low rank tensor $X$. 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 4 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 $m_0$ denotes the maximal number of measurements for which not even one out of 200 generated tensors is recovered and $m_1$ denotes the minimal number of measurements for which all 200 tensors are recovered. The fifth column in Table 4 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. Alternatively, the SDPNAL+ Matlab toolbox (version 0.5 beta) for semidefinite programming [61, 63] allows to perform low rank tensor recovery via $\theta_1$-norm minimization for even higher-dimensional tensors. For example, with $m = 95$ measurement we managed to recover all rank-one $9 \times 9 \times 9$ tensors out of 200 (each simulation taking about 5min). Similarly, rank-one $11 \times 11 \times 11$ tensors are recovered from $m = 125$ measurements with one simulation lasting about 50min. Due to these large computation times, more elaborate numerical experiments have not been conducted in these scenarios. We remark that no attempt of accelerating the optimization algorithm has been made. This task is left for future research.

Table 4. Numerical results for low rank tensor recovery in $\mathbb{R}^{n_1 \times n_2 \times n_3}$.

| $n_1 \times n_2 \times n_3$ | rank | $m_0$ | $m_1$ | $n_1n_2n_3$ | cpu (sec) |
|-----------------------------|------|-------|-------|--------------|-----------|
| $2 \times 2 \times 3$       | 1    | 4     | 12    | 12           | 0.2       |
| $3 \times 3 \times 3$       | 1    | 6     | 19    | 27           | 0.37      |
| $3 \times 4 \times 5$       | 1    | 11    | 30    | 60           | 6.66      |
| $4 \times 4 \times 4$       | 1    | 11    | 32    | 64           | 7.28      |
| $4 \times 5 \times 6$       | 1    | 18    | 42    | 120          | 129.48    |
| $5 \times 5 \times 5$       | 1    | 18    | 43    | 125          | 138.90    |
| $3 \times 4 \times 5$       | 2    | 27    | 56    | 60           | 7.55      |
| $4 \times 4 \times 4$       | 2    | 26    | 56    | 64           | 8.65      |
| $4 \times 5 \times 6$       | 2    | 41    | 85    | 120          | 192.58    |

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 tensor space. Therefore, low rank tensor recovery via $\theta_1$-minimization seems to be a promising approach. Of course, it remains to investigate the recovery performance theoretically.

Figures 4 and 5 present the numerical results for low rank tensor recovery via $\theta_1$-norm minimization for Gaussian measurement maps, conducted with the SDPNAL+ toolbox. For fixed tensor dimensions $n \times n \times n$, fixed tensor rank $r$, and fixed number $m$ of measurements 50 simulations are performed. We say that recovery is successful if the element-wise reconstruction error is smaller than $10^{-5}$. Figures 4a, 5a and 1b, 2b, 3b present experiments for rank-one and rank-two tensors, respectively. The vertical axis in all three figures represents the empirical success rate. In Figure 1 the horizontal axis represents the relative number of measurements,
The computation is implemented in MATLAB R2016a, on an Acer Laptop with CPU@1.90GHz and RAM 4GB. With $C > 0$ such that $m \geq 3C r n$, for an universal constant $C$, Figure 2 suggests that the constant $C$ depends on the size of the tensor. In particular, it seems to grow slightly with $n$ (although it is still possible that there exists $C > 0$ such that $m \geq 3C r n$ would always be enough for the recovery). With $C = 3.3$ we would always be able to recover a low rank tensor of size $n \times n \times n$ with $n \leq 7$. The horizontal axis in Figure 2 represents the number $m/(3nr \cdot \log(n))$. The figure suggests that with the number of measurements $m \geq 6nr \cdot \log(n)$ we would always be able to recover a low rank tensor and therefore it may be possible that a logarithmic factor is necessary.

The computation is implemented in MATLAB R2016a, on an Acer Laptop with CPU@1.90GHz and RAM 4GB.

We remark that we have used standard MATLAB packages for convex optimization to perform the numerical experiments. To obtain better performance, new optimization methods should be developed specifically to solve our optimization problem, or more generally, to solve the sum-of-squares polynomial problems. We
Figure 3. Recovery of rank-1 and rank-2 tensors via $\theta_1$-norm minimization.

expect this to be possible and the resulting algorithms to give much better performance results since we have shown that in the matrix scenario all theta norms correspond to the matrix nuclear norm. The state-of-the-art algorithms developed for the matrix scenario can compute the matrix nuclear norm and can solve the matrix nuclear norm minimization problem for matrices of large dimensions. The theory developed in this paper together with the first numerical results should encourage the development into this direction.

Appendix A. Monomial orderings and Gröbner bases

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

**Definition A.1.** For $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)$, $\beta = (\beta_1, \beta_2, \ldots, \beta_n) \in \mathbb{Z}_+^n$, we write $x^\alpha >_{\text{grevlex}} x^\beta$ (or $\alpha >_{\text{grevlex}} \beta$) if $|\alpha| > |\beta|$ or $|\alpha| = |\beta|$ and the rightmost nonzero entry of $\alpha - \beta$ 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 [14, 15].

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 [5].

**Definition A.2** (Gröbner basis). For a fixed monomial order, a basis $\mathcal{G} = \{g_1, \ldots, g_s\}$ of a polynomial ideal $J \subset \mathbb{R}[x]$ is a Gröbner basis (or standard basis) if for all $f \in \mathbb{R}[x]$ there exist a unique $r \in \mathbb{R}[x]$ 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 $\text{LM}(g_1), \text{LM}(g_2), \ldots, \text{LM}(g_s)$.

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

**Definition A.3.** The reduced Gröbner basis for a polynomial ideal $J \subset \mathbb{R}[x]$ is a Gröbner basis $\mathcal{G} = \{g_1, g_2, \ldots, g_s\}$ for $J$ such that

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

\[ \text{(A) Recovery of rank-1 tensors} \]

\[ \text{(B) Recovery of rank-2 tensors} \]
In other words, a Gröbner basis $G$ is the reduced Gröbner basis if for all $i \in [s]$ the polynomial $g_i \in G$ is monic (i.e., $\text{LC}(g_i) = 1$) and the leading monomial $\text{LM}(g_i)$ 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 A.2 equals zero. Gröbner bases are also one of the main computational tools in solving systems of polynomial equations \[15\].

With $\overline{f}^G$ 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 \( (f_1, f_2, \ldots, f_k) \), then we can regard $F$ as a set without any particular order by Definition A.2, or in other words, the result of the division algorithm does not depend on the order of the polynomials. Therefore, $\overline{f}^G = r$ in Definition A.2.

The following result follows directly from Definition A.2 and the polynomial division algorithm \[15\].

**Corollary A.4.** Fix a monomial ordering and let $G = \{g_1, g_2, \ldots, g_s\} \subset \mathbb{R}[x]$ be a Gröbner basis of a polynomial ideal $J$. A polynomial $f \in \mathbb{R}[x]$ is in the ideal $J$ if it can be written in the form $f = a_1g_1 + a_2g_2 + \ldots + a_sg_s$, where $a_i \in \mathbb{R}[x]$, for all $i \in [s]$, s.t. whenever $a_ig_i \neq 0$ we have

$$\text{multideg}(f) \geq \text{multideg}(a_ig_i).$$

**Definition A.5.** Fix a monomial order and let $G = \{g_1, g_2, \ldots, g_s\} \subset \mathbb{R}[x]$. Given $f \in \mathbb{R}[x]$, we say that $f$ reduces to zero modulo $G$ and write

$$f \rightarrow_G 0$$

if it can be written in the form $f = a_1g_1 + a_2g_2 + \ldots + a_sg_s$ with $a_i \in \mathbb{R}[x]$ for all $i \in [k]$ s.t. whenever $a_ig_i \neq 0$ we have $\text{multideg}(f) \geq \text{multideg}(a_ig_i)$.

Assume that $G$ 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 $G$. In other words, for a Gröbner basis $G$,

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

The Gröbner basis of a polynomial ideal always exists and can be computed in a finite number of steps via Buchberger’s algorithm \[5\] \[14\] \[15\].

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 A.6.** Let $f, g \in \mathbb{R}[x]$ be a non-zero polynomials.

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

2. The $S$-polynomial of $f$ and $g$ is the combination

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

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

**Theorem A.7 (Buchberger’s criterion).** A basis $G = \{g_1, g_2, \ldots, g_s\}$ for a polynomial ideal $J \subset \mathbb{R}[x]$ is a Gröbner basis if and only if $S(g_i, g_j) \rightarrow_G 0$ for all $i \neq j$.  

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

**Proposition A.8.** Given a finite set $G \subset \mathbb{R}[x]$, suppose that the leading monomials of $f, g \in G$ are relatively prime, i.e.,

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

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

Therefore, to prove that the set $G \subset \mathbb{R}[x]$ is a Gröbner basis, it is enough to show that $S(g_i, g_j) \rightarrow_{G} 0$ 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 and support by the Hausdorff Research Institute for Mathematics, Bonn through the trimester program Mathematics of Signal Processing. Most of the research was done while Ž. S. was a PhD student at the University of Bonn and employed at RWTH Aachen University.

**References**

[1] R. Bhatia. *Matrix Analysis*. Graduate Texts in Mathematics. 169. Springer, 1996.

[2] G. Blekherman, P. A. Parrilo, and R. R. Thomas. *Semidefinite optimization and convex algebraic geometry*. SIAM, 2013.

[3] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge Univ. Press, 2004.

[4] J.-L. Brylinski. Algebraic measures of entanglement. In G. Chen and R. K. Brylinski, editors, *Mathematics of Quantum Computation*. CRC, Boca Raton, FL, 2002.

[5] 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.

[6] 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.

[7] E. J. Candès and B. Recht. Exact matrix completion via convex optimization. *Found. Comput. Math.*, 9(6):717–772, 2009.

[8] 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.

[9] E. J. Candès and T. Tao. The power of matrix completion: near-optimal convex relaxation. *IEEE Trans. Information Theory*, 56(5):2053–2080, 2010.

[10] 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.

[11] D. Cartwright, D. Erman, and L. Oeding. Secant varieties of $\mathbb{P}^2 \times \mathbb{P}^n$ embedded by $\mathcal{O}(1,2)$. *Journal of the London Mathematical Society*, 85(1):121–141, 2012.
[12] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. Willsky. The convex geometry of linear inverse problems. *Found. Comput. Math.*, 12(6):805–849, 2012.

[13] Y. Chen, S. Bhojanapalli, S. Sanghavi, and R. Ward. Completing any low-rank matrix, provably. *Journal of Machine Learning Research*, 16:2999–3034, 2015.

[14] D. Cox, J. Little, and D. O’Shea. *Using algebraic geometry*, volume 185 of *Graduate Texts in Mathematics*. Springer, New York, Second edition, 2005.

[15] D. Cox, J. Little, and D. O’Shea. *Ideals, varieties, and algorithms*. Undergraduate Texts in Mathematics. Springer, New York, Third edition, 2007.

[16] C. Da Silva and F. J. Herrmann. Hierarchical tucker tensor optimization-applications to tensor completion. In *SAMPTA 2013*, pages 384–387, 2013.

[17] V. De Silva and L.-H. Lim. Tensor rank and ill-posedness of the best low-rank approximation problem. *SIAM Journal on Matrix Analysis and Applications*, 30(3):1084–1127, 2008.

[18] A. Defant and K. Floret. *Tensor Norms and Operator Ideals*. North-Holland Mathematics Studies. Elsevier Science, 1992.

[19] D. L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.

[20] M. F. Duarte and R. G. Baraniuk. Kronecker compressive sensing. *Image Proc., IEEE Trans. on*, 2011.

[21] M. Fazel. *Matrix rank minimization with applications*. PhD thesis, 2002.

[22] S. Foucart and H. Rauhut. *A Mathematical Introduction to Compressive Sensing*. Applied and Numerical Harmonic Analysis. Birkhäuser, 2013.

[23] S. Friedland and L.-H. Lim. Nuclear Norm of Higher-Order Tensors. *Mathematics of Computation, accepted for publication*, 2016.

[24] S. Gandy, B. Recht, and I. Yamada. Tensor completion and low-n-rank tensor recovery via convex optimization. *Inverse Problems*, 27(2):19pp, 2011.

[25] J. Gouveia, M. Laurent, P. A. Parrilo, and R. R. Thomas. A new semidefinite programming hierarchy for cycles in binary matroids and cuts in graphs. *Mathematical Programming*, pages 1–23, 2009.

[26] J. Gouveia, P. A. Parrilo, and R. R. Thomas. Theta bodies for polynomial ideals. *SIAM J. Optim.*, 20(4):2097–2118, 2010.

[27] F. Grande and R. Sanyal. Theta rank, levelness, and matroid minors. *Journal of Combinatorial Theory, Series B*, pages –, 2016.

[28] L. Grasedyck. Hierarchical singular value decomposition of tensors. *SIAM. J. Matrix Anal. & Appl*, 31:2029, 2010.

[29] 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.

[30] D. Gross. Recovering low-rank matrices from few coefficients in any basis. *IEEE Trans. Inform. Theory*, 57(3):1548–1566, 2011.

[31] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert. Quantum state tomography via compressed sensing. *Phys. Rev. Lett.*, 105:150401, 2010.

[32] W. Hackbusch. *Tensor spaces and numerical tensor calculus*. Springer, 2012.

[33] T. Hibi. Distributive lattices, affine semigroup rings and algebras with straightening laws. *Commutative algebra and combinatorics, US-Jap. joint Semin., Kyoto/Jap. 1985*, Advanced Studies in Pure Mathematics 11, 93-109 (1987), 1987.

[34] C. J. Hillar and L.-H. Lim. Most tensor problems are NP-hard. *J. ACM*, 60(6):45:1–45:39, 2013.

[35] F. L. Hitchcock. The expression of a tensor or a polyadic as a sum of products. *Journal of Mathematics and Physics*, 6(1-4):164–189, 1927.
[36] F. L. Hitchcock. Multiple invariants and generalized rank of a p-way matrix or tensor. *J. Math. Phys.*, 7(1):39–79, 1927.

[37] J. Hästad. Tensor rank is NP-complete. *J. Algorithms*, 11(4):644–654, 1990.

[38] B. Huang, C. Mu, D. Goldfarb, and J. Wright. Provable models for robust low-rank tensor recovery. *Pac. J. Optim.*, 11(2):339–364, 2015.

[39] L. Karlsson, D. Kressner, and A. Uschmajew. Parallel algorithms for tensor completion in the CP format. *Parallel Comput.*, 2015. In press.

[40] 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.

[41] D. Kressner, M. Steinlechner, and B. Vandereycken. Low-rank tensor completion by riemannian optimization. *BIT Numerical Mathematics*, 54(2):447–468, 2014.

[42] R. Kueng, H. Rauhut, and U. Terstiege. Low rank matrix recovery from rank one measurements. *Appl. Comput. Harmon. Anal.*, to appear, 2014.

[43] J. M. Landsberg. *Tensors: Geometry and Applications*. Graduate studies in mathematics. American Mathematical Society, 2011.

[44] J. Lasserre. *Moments, positive polynomials and their applications*, volume 1 of *Imperial College Press Optimization Series*. Imperial College Press, London, 2010.

[45] J. Liu, P. Musialski, P. Wonka, and J. Ye. Tensor completion for estimating missing values in visual data. In *ICCV*, 2009.

[46] 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.

[47] L. Lovász. On the Shannon capacity of a graph. *IEEE Trans. Inform. Theory*, 25(1):1–7, 1979.

[48] I. V. Oseledets. Tensor-train decomposition. *SIAM J. Sci. Comput.*, 33(5):2295–2317, 2011.

[49] S. Oymak, A. Jalali, M. Fazel, Y. C. Eldar, and B. Hassibi. Simultaneously structured models with application to sparse and low-rank matrices. *IEEE Transactions on Information Theory*, 61(5):2886–2908, 2015.

[50] N. Parikh and S. Boyd. Proximal algorithms. *Foundations and Trends in Optimization*, 1(3):123–231, 2013.

[51] H. Rauhut, R. Schneider, and Ž. Stojanac. Tensor tensor recovery via iterative hard thresholding. In *Proc. SampTA 2013*, 2013.

[52] 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.

[53] H. Rauhut, R. Schneider, and Ž. Stojanac. Low rank tensor recovery via iterative hard thresholding. *arXiv preprint arXiv:1602.05217*, 2016.

[54] 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.

[55] B. Romera-Paredes, H. Aung, N. Bianchi-Berthouze, and M. Pontil. Multilinear multitask learning. *J. Mach. Learn. Res.*, 28(3):1444–1452, 2013.

[56] R. A. Ryan. *Introduction to Tensor Products of Banach Spaces*. Celtic Studies. Springer, 2002.

[57] Ž. Stojanac. *Low-rank Tensor Recovery*. PhD thesis, Universitä Bonn, 2016.

[58] K. Toh and S. Yun. An accelerated proximal gradient algorithm for nuclear norm regularized least squares problems. *Pac. J. Optim.*, 6:615–640, 2010.

[59] R. Tomioka, K. Hayashi, and H. Kashima. Estimation of low-rank tensors via convex optimization. *arXiv preprint arXiv:1010.0789*, 2010.
[60] Y.-C. Wong. *Schwartz spaces, nuclear spaces, and tensor products*. Lecture notes in mathematics. Springer-Verlag, 1979.

[61] L. Q. Yang, D. F. Sun, and K. C. Toh. SDPNAL+: a majorized semismooth Newton-CG augmented Lagrangian method for semidefinite programming with nonnegative constraints. *Mathematical Programming Computation*, 7(3):331–366, 2015.

[62] M. Yuan and C.-H. Zhang. On tensor completion via nuclear norm minimization. *Foundations of Computational Mathematics*, 16(4):1031–1068, 2016.

[63] X. Y. Zhao, D. F. Sun, and K. C. Toh. A Newton-CG Augmented Lagrangian Method for Semidefinite Programming. *SIAM J. Optimization*, 20(4):1737–1765, 2010.

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

RWTH Aachen University, University of Cologne, Institute for Theoretical Physics, Zülpicher Straße 77, 50937 Germany  
E-mail address: stojanac@thp.uni-koeln.de