



Tensor theta norms and low rank recovery

Holger Rauhut¹ · Željka Stojanac¹

Received: 7 July 2019 / Accepted: 8 October 2020 / Published online: 27 November 2020
© The Author(s) 2020

Abstract

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

Keywords Low rank tensor recovery · Tensor nuclear norm · Theta bodies · Compressive sensing · Semidefinite programming · Convex relaxation · Polynomial ideals · Gröbner bases

✉ Holger Rauhut
rauhut@mathc.rwth-aachen.com

Mathematics Subject Classification 2010 13P10 · 15A69 · 15A60 · 52A41 · 90C22 · 90C25 · 94A20

1 Introduction and motivation

Compressive sensing predicts that sparse vectors can be recovered from underdetermined linear measurements via efficient methods such as ℓ_1 -minimization [10, 20, 23]. 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 [22, 55]. 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 [6, 7, 31, 32, 43, 55].

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 [46], reflectance data recovery [46] (e.g., for use in photo-realistic raytracers), machine learning [56], and seismic data processing [41]. Several approaches have already been introduced [25, 39, 46, 52, 53], but unfortunately, so far, for none of them a completely satisfactory theory is available. Either the method is not tractable [63], or no (complete) rigorous recovery results quantifying the minimal number of measurements are available [17, 25, 40, 42, 46, 52, 53], or the available bounds are highly nonoptimal [21, 39, 47]. For instance, the computation (and therefore, also the minimization) of the tensor nuclear norm ([19, 57, 61]) for higher order tensors is in general NP-hard [24]—nevertheless, some recovery results for tensor completion via nuclear norm minimization are available in [63]. Moreover, versions of iterative hard thresholding for various tensor formats have been introduced [52, 53]. 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 [52–54]. 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 [53, 54].

This article introduces a new approach for tensor recovery based on convex relaxation, initially suggested in slightly different form (but not worked out) in [12]. 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, 27, 48] which is similar to the better known Lasserre relaxations [45]. 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- θ_k -norm balls converges asymptotically to the unit tensor nuclear norm ball. The θ_k -norm as well as its minimization subject to an affine constraint can be computed via semidefinite optimization, and also the minimization of the θ_k -norm subject to a linear constraint is a semidefinite program (SDP), whose solution can be computed in polynomial time—the complexity growing with k .

The tensor nuclear norm may be defined for both base fields $\mathbb{R}$ and $\mathbb{C}$. In general the resulting norms may differ for a real tensor. In this article, we restrict to relaxations of the real tensor nuclear norm because the concept of theta bodies is based on *real* algebraic geometry and not well-defined for the complex case.

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 (a certain 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.

Our investigations have been strongly motivated by [12], where theta bodies have first been suggested for low rank tensor recovery. The approach in [12] has not been worked out in detail, however. It suggests a slightly different polynomial ideal that requires additional auxiliary variables. The corresponding Gröbner basis and, hence, also the theta basis, become much more complicated (see also Remark 2). This would lead to very technical computations on the theoretical side and to less efficient algorithms on the practical side.

We show that for the matrix case (tensors of order 2), our relaxation 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.

The derivation of the semidefinite program for calculating the θ_k -norm requires to compute the so-called theta basis of the related polynomial ideal which in turn needs the reduced Gröbner basis. We prove the somewhat surprising fact that the Gröbner basis is given by the generating set defining the polynomial ideal, i.e., the order two minors and the polynomial related to the Frobenius norm. This is one of the core results of this paper. Its proof is somewhat technical (and therefore we separate the simpler order three case from the case of tensors of general order d), but it allows us

to explicitly compute the theta basis and the so-called moment matrix, which finally defines the semidefinite program.

We present numerical experiments which show that θ_1 -norm minimization successfully recovers tensors of low rank from few random linear measurements. We remark that we use a standard semidefinite solver which limits the size of tensors as computation time becomes too large (despite formally being polynomial) for tensors whose size is of the order $10 \times 10 \times 10$, say. This may seem a severe limitation, but we emphasize that the focus of this paper is a first investigation of the tensor θ_k -norms with a derivation of the corresponding semidefinite programs and first (promising) numerical tests on the recovery performance. We expect that specialized algorithms for θ_k -norm minimization, for instance based on proximal splitting methods [14] such as ADMM, may lead to significantly increased computation speed with respect to standard semidefinite solvers. A second main motivation of our work is that the θ_k -norm minimization approach seems like a promising polynomially tractable approach that allows for a theoretical analysis of the required number of random linear measurement ensuring recovery—improving over presently available bounds for tractable algorithms. As outlined above, optimal estimates of the required number of measurements are presently available only for tensor recovery approaches that are NP-hard. Unfortunately, such a theoretical analysis is still missing for θ_k -norm minimization, but will be the subject of future work. In this sense, the present article may be seen as a contribution that hopefully paves the way for a better understanding of the theory of low rank tensor recovery.

Contributions We summarize the main contributions of this article below.

- We show that the θ_k -norm reduces to nuclear norm in the matrix case for all $k \in \mathbb{N}$. This fact suggest that the θ_k -norms are natural generalizations of the matrix nuclear norm to the tensor case.
- We provide semidefinite programs for the calculation of the θ_k -norms in the case of general tensors of order $d \geq 3$. We present numerical experiments for low rank tensor recovery from a small number of random Gaussian linear measurements which show that our approach is successful in practice.
- The derivation of the semidefinite programs requires to compute a moment matrix based on a theta basis of the vector space of real polynomials modulo the ideal. The computation of the theta basis in turn needs a reduced Gröbner basis of the polynomial ideal whose real algebraic variety corresponds to the (canonical) rank one, unit norm tensors. We prove the remarkable fact of potential independent interest that the generating set of minors of order 2 and the squared Frobenius norm minus 1 is already a Gröbner with respect to the graded reverse lexicographic (grevlex) ordering (Section 4.1 for third-order tensors, and Section 4.2 for general d th-order tensors).
- We show in addition that no matter which notion of tensor rank (canonical, TT, HOSVD) we consider, the polynomial ideal generated by the rank one (in the corresponding notion), unit norm tensors are all the same. As a consequence, the θ_k -norms corresponding to the different notions will all coincide (Section 4.2).
- Due to the fact that the theta norms are built from the polynomial ideal whose real algebraic variety contains all rank-one unit norm tensors, it is a natural

question to ask whether the resulting θ_k -norms coincide with the weighted sum of the nuclear norms of the matricizations. In Remark 3 (Section 4.1) we show that this is not the case at least for the largest relaxation, i.e., for the θ_1 -norm.

- We prove that the sequence of θ_k -norms convergence asymptotically to the (real) tensor nuclear norm as $k \rightarrow \infty$ (Section 5).

The last point should be seen as a rather theoretical result because in practice one would rather choose $k = 1$ or $k = 2$ due to computational constraints. Therefore, one cannot easily transfer theoretical results for tensor nuclear norm minimization to θ_k -norm minimization, but one rather requires a direct analysis of our approach which is postponed to future contributions.

1.1 Low rank matrix recovery

Before passing to tensor recovery, we recall some basics on matrix recovery. Let $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2}$ of rank at most $r \ll \min\{n_1, n_2\}$, and suppose we are given linear measurements $\mathbf{y} = \mathcal{A}(\mathbf{X})$, where $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \rightarrow \mathbb{R}^m$ is a linear map with $m \ll n_1 n_2$. Reconstructing $\mathbf{X}$ from $\mathbf{y}$ amounts to solving an underdetermined linear system. Unfortunately, the rank minimization problem of computing the minimizer of

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

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

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

has been suggested [22, 55], where the nuclear norm $\|\mathbf{Z}\|_* = \sum_j \sigma_j(\mathbf{Z})$ is the sum of the singular values of $\mathbf{Z}$. This problem can be solved efficiently by various methods [3]. For instance, it can be reformulated as a semidefinite program [22], but splitting methods may be more efficient [14, 51, 59].

A by-now standard result [6, 12] states that a matrix $\mathbf{X}$ of rank r can be stably recovered from $\mathbf{y} = \mathcal{A}(\mathbf{X})$, where $\mathcal{A}$ is a Gaussian measurement map, via nuclear norm minimization (1) with probability at least $1 - e^{-cm}$ provided that

$$m \geq Cr(n_1 + n_2), \quad (2)$$

where the constants $c, C > 0$ are universal. Other interesting measurement maps (matrix completion and rank-one measurements) have been studied in [7–9, 13, 31, 43].

1.2 Tensor recovery

An order- d tensor (or mode- d -tensor) is an element $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ indexed by $[n_1] \times [n_2] \times \cdots \times [n_d]$. Of course, the case $d = 2$ corresponds to matrices. For $d \geq 3$, several notions and computational tasks become much more involved than for the matrix case. Already the notion of rank requires some clarification, and in fact, several different definitions are available (see, for instance, [30, 36, 37, 44]). We will mainly work with the canonical rank or CP-rank in the following. A d th-order tensor

$\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ is of rank one if there exist vectors $\mathbf{u}^1 \in \mathbb{R}^{n_1}$, $\mathbf{u}^2 \in \mathbb{R}^{n_2}$, $\dots$, $\mathbf{u}^d \in \mathbb{R}^{n_d}$ such that $\mathbf{X} = \mathbf{u}^1 \otimes \mathbf{u}^2 \otimes \cdots \otimes \mathbf{u}^d$ or elementwise

$$X_{i_1 i_2 \dots i_d} = u_{i_1}^1 u_{i_2}^2 \cdots u_{i_d}^d.$$

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

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

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

where the nuclear norm is defined as

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

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 [24, 35, 38]). Let us nevertheless mention that some theoretical results for tensor recovery via nuclear norm minimization are contained in [63].

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, 18, 24]). 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 $\mathbf{X} \in \mathbb{R}^{n \times n \times n}$ into a low rank matrix $\tilde{\mathbf{X}} \in \mathbb{R}^{n \times n^2}$ and simply apply concepts from matrix recovery. However, the bound (2) on the required number of measurements then reads

$$m \geq C r n^2. \quad (3)$$

Moreover, it has been suggested in [25, 46, 60] 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 [50], 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.

A version of the restricted isometry property for certain tensor formats in [54] is satisfied for

$$m \geq Cr^2n \quad (4)$$

Gaussian random measurements with high probability—precisely, this bound uses the tensor train format [49]. (Possibly, the term r^2 may even be lowered to r when using the “right” tensor format.) Unfortunately, up to the authors knowledge, it is open to show that an efficient (polynomial time) algorithm can recover rank r tensors if the restricted isometry property is satisfied. Only partial results are known [53, 54]: a tensor iterative hard thresholding algorithm is shown to converge to the original rank r tensor if on top of the restricted isometry property a certain inequality is satisfied for the approximate projection of each iterate onto the rank r tensors. Unfortunately, that inequality cannot be guaranteed for the approximate projection and also cannot be checked throughout the iterations. The exact projection would satisfy it, but is NP-hard to compute, which is the reason why one resorts to an efficient approximate projection. (Given the empirical success of the algorithm, it seems that the inequality usually holds at least starting from a certain iteration.) A local convergence result for tensor iterative hard thresholding has been given in [53], but one cannot guarantee that the iterates get close enough to the original low rank tensor ensuring convergence to the original tensor by the local result.

In any case, considering that the bound (4) for an RIP adapted to certain tensor formats is significantly better than (3) 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. It may be expected that the bound on the minimal number of measurements needed for low rank tensor recovery via tensor nuclear norm minimization is optimal. We conjecture that such optimal bound is of the form $m \geq Crn$ or possibly $m \geq Crn \log(n)$. (Our numerical experiments suggest that at least the latter is true, see Fig. 3.) We note that it has been shown in [63] that tensor completion via tensor nuclear norm minimization is successful in recovering (incoherent) $n \times n \times n$ tensors of rank r if $m \geq C\sqrt{r}(n \log(n))^{3/2}$, which is slightly worse than the conjectured bound (in particular, $\sqrt{rn}$ instead of r). This deficiency may be due to the fact that tensor completion is harder than recovery from Gaussian random matrices or that the proof given in [63] does not give the optimal bound (or both). In any case, the drawback of tensor nuclear norm minimization is that the tensor nuclear norm is NP-hard to compute so that this approach is intractable. In fact, [63] only gives a theoretical analysis and no algorithm (not even a heuristic one) for solving tensor nuclear norm minimization problems.

To overcome this difficulty, we introduce what we call the tensor θ_k -norms in this paper—new tensor norms which can be computed via semidefinite programming. These norms are tightly related to the tensor nuclear norm. That is, the unit θ_k -norm balls (which are defined for $k \in \mathbb{N}$) satisfy

$$\begin{aligned} \{\mathbf{X} : \|\mathbf{X}\|_{\theta_1} \leq 1\} &\supseteq \cdots \supseteq \{\mathbf{X} : \|\mathbf{X}\|_{\theta_k} \leq 1\} \supseteq \{\mathbf{X} : \|\mathbf{X}\|_{\theta_{k+1}} \leq 1\} \\ &\supseteq \cdots \supseteq \{\mathbf{X} : \|\mathbf{X}\|_* \leq 1\}. \end{aligned}$$

In particular, we show that in the matrix scenario all θ_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- θ_k -norm balls converges asymptotically to the unit tensor nuclear norm ball. Next, we provide numerical experiments on low rank tensor recovery via θ_1 -norm minimization. We provide numerical experiments for θ_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 θ_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 $\mathcal{S}$ is denoted by $|\mathcal{S}|$.

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

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 $\|\mathbf{X}\|_F = \left(\sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \cdots \sum_{i_d=1}^{n_d} X_{i_1 i_2 \dots i_d}^2 \right)^{1/2}$. The vectorization of a tensor $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ is denoted by $(\mathbf{X}) \in \mathbb{R}^{n_1 n_2 \dots n_d}$. 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. For a tensor $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ and an ordered subset $\mathcal{S} \subseteq [d]$, an $\mathcal{S}$ -matricization $\mathbf{X}^{\mathcal{S}} \in \mathbb{R}^{\prod_{k \in \mathcal{S}} n_k \times \prod_{\ell \in \mathcal{S}^c} n_\ell}$ is defined as $X_{(i_k)_{k \in \mathcal{S}}, (i_\ell)_{\ell \in \mathcal{S}^c}}^{\mathcal{S}} = X_{i_1 i_2 \dots i_d}$, i.e., the indices in the set $\mathcal{S}$ define the rows of the matrix $\mathbf{X}^{\mathcal{S}}$ and the indices in the set $\mathcal{S}^c = [d] \setminus \mathcal{S}$ define the columns. For a singleton set $\mathcal{S} = \{i\}$, for $i \in [d]$, we call the $\mathcal{S}$ -matricization the i th unfolding.

For a tensor $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ of order d , we write $\mathbf{X}(:, :, \dots, :, k)$ for the order $(d-1)$ subtensor in $\mathbb{R}^{n_1 \times n_2 \times \cdots \times n_{d-1}}$ obtained by fixing the last index α_d to k . Instead of writing $x_{\alpha_1 \alpha_2 \dots \alpha_d} x_{\beta_1 \beta_2 \dots \beta_d}$, we often use the simpler notation $x_{\alpha} x_{\beta}$. We will use the grevlex ordering of monomials: $x_{11\dots 11} > x_{11\dots 12} > \cdots > x_{11\dots 1n_d} > x_{111\dots 21} > \cdots > x_{n_1 n_2 \dots n_d}$.

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 θ_k -norms and show that they all coincide with the matrix nuclear norm. The case of 2×2 -matrices is described in detail. In Section 4 we pass to the tensor case and discuss first the case of order-three tensors. We introduce a suitable polynomial ideal, provide its reduced Gröbner basis and define the corresponding θ_k -norms. We additionally show that considering matricizations corresponding to the TT-format will lead to the same polynomial ideal and thus to the same θ_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 θ_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 θ_k -norms. In Section 5 we discuss the convergence of the sequence of the unit- θ_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 θ_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 discusses some background from computer algebra (monomial orderings and Gröbner bases) that is required throughout the main body of the article.

2 Theta bodies

As outlined above, we will introduce new tensor norms as relaxations of the nuclear norm in order to come up with a new convex optimization approach for low rank tensor recovery. Our approach builds on theta bodies, a recent concept from computational algebraic geometry, which is similar to Lasserre relaxations [45]. In order to introduce it, we first discuss the necessary basics from computational commutative algebra. For more information, we refer to [15, 16] and to the Appendix.

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

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

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

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

By Hilbert's basis theorem [16] every polynomial ideal in $\mathbb{R}[\mathbf{x}]$ has a finite generating set. Thus, we may assume that J is generated by a set $\mathcal{F} = \{f_1, f_2, \dots, f_k\}$ of polynomials in $\mathbb{R}[x]$ and write

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

Its real algebraic variety is the set

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

Throughout the paper, $\mathbb{R}[\mathbf{x}]_k$ denotes the set of polynomials of degree at most k . A degree one polynomial is also called linear polynomial. A very useful certificate for positivity of polynomials is contained in the following definition [27].

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

A special case of theta bodies was first introduced by Lovász in [48] and in full generality they appeared in [27]. Later, they have been analyzed in [26, 28]. The definitions and theorems in the remainder of the section are taken from [27].

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

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

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

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

$$TH_1(J) \supseteq TH_2(J) \supseteq \dots \supseteq \text{conv}(v_{\mathbb{R}}(J)). \quad (5)$$

The theta body sequence of J can converge (finitely or asymptotically), if at all, only to $\text{conv}(v_{\mathbb{R}}(J))$. More on guarantees on convergence can be found in [27, 28]. However, to our knowledge, none of the existing guarantees applies to the cases discussed below.

Given any polynomial, it is possible to check whether it is k -sos mod J using a Gröbner basis and semidefinite programming. However, using this definition in practice requires knowledge of all linear polynomials (possibly infinitely many) that are k -sos mod J . To overcome this difficulty, we need an alternative description of $TH_k(J)$ 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 $\mathbb{R}[x'] = \mathbb{R}[x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_n]$. Therefore, $\mathbb{R}[x]_1/J \cong \mathbb{R}[x]_1$ and $\{1 + J, x_1 + J, \dots, x_n + J\}$ can be completed to a basis $\mathcal{B}$ 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 $\mathcal{B} = \{f_i + J\}$ 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 $\mathcal{B} = \{f_i + J\}$ is ordered so that $f_{i+1} > f_i$, where $>$ is a fixed monomial ordering. Further, we define the set $\mathcal{B}_k$

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

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

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

As in [2, 27] we consider only monomial bases $\mathcal{B}$ of $\mathbb{R}[\mathbf{x}]/J$, i.e., bases $\mathcal{B}$ such that f_i is a monomial, for all $f_i + J \in \mathcal{B}$.

For determining a θ -basis, we first need to compute the reduced Gröbner basis $\mathcal{G}$ of the ideal J (see Definitions 8 and 9). The set $\mathcal{B}$ will satisfy the second property in the definition of the theta basis if the reduced Gröbner basis is with respect to an ordering which first compares the total degree. Therefore, throughout the paper we use the graded reverse monomial ordering (Definition 7) or simply grevlex ordering, although also the graded lexicographic ordering would be appropriate.

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

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

where $\mathcal{G} = \{g_1, g_2, \dots, g_s\}$ is the reduced Gröbner basis of the ideal J . In other words, a set $\mathcal{B} = \{f_0 + J, f_1 + J, \dots\}$ will be a θ -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 $\{LT(g_i) : i \in [s]\}$.

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

The theta bodies can be characterized via the combinatorial moment matrix as stated in the next result from [27], which will be the basis for computing and minimization the new tensor norm introduced below via semidefinite programming.

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

Theorem 1 *The k th theta body of J , $TH_k(J)$, is the closure of*

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

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

Algorithm 1 shows a step-by-step procedure for computing $TH_k(J)$.

Algorithm 1 Algorithm for computing $TH_k(J)$.

Input: An ideal $J \in \mathbb{R}[\mathbf{x}] = \mathbb{R}[x_1, x_2, \dots, x_n]$.

Compute the reduced Gröbner basis for the ideal J

Compute a θ -basis $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2 \cup \dots = \{f_0 + J, f_1 + J, \dots\}$ of $\mathbb{R}[\mathbf{x}]/J$ (see Definition 3)

Compute the combinatorial moment matrix $\mathbf{M}_{\mathcal{B}_k}(\mathbf{y})$:

1. $[\mathbf{x}]_{\mathcal{B}_k} = \{\text{all elements of } \mathcal{B}_k \text{ in order}\}$
2. $(\mathbf{X}_{\mathcal{B}_k})_{i,j} = ([\mathbf{x}]_{\mathcal{B}_k} [\mathbf{x}]_{\mathcal{B}_k}^T)_{i,j} = f_i f_j + J = \sum_{f_l + J \in \mathcal{B}_{2k}} \lambda_{i,j}^l (f_l + J)$
3. $[\mathbf{M}_{\mathcal{B}_k}(\mathbf{y})]_{i,j} = \sum_{f_l + J \in \mathcal{B}_{2k}} \lambda_{i,j}^l y_l$

Output: $TH_k(J)$ is the closure of

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

3 The matrix case

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

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

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

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

$$v_{\mathbb{R}}(J) = \{\mathbf{X} \in \mathbb{R}^{m \times n} : \|\mathbf{X}\|_F = 1, \text{rank}(\mathbf{X}) = 1\}. \quad (6)$$

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

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

$$g(\mathbf{x}) = \sum_{i=1}^m \sum_{j=1}^n x_{ij}^2 - 1 \text{ and } f_{ijkl}(\mathbf{x}) = x_{il}x_{kj} - x_{ij}x_{kl} \\ \text{for } 1 \leq i < k \leq m, 1 \leq j < l \leq n. \quad (7)$$

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

$$\mathbf{X} \in \mathcal{R} := \{\mathbf{X} : g(\mathbf{X}) = 0 \text{ and } f_{ijkl}(\mathbf{X}) = 0 \text{ for all } i < k, j < l\}. \quad (8)$$

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

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

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

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

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

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

$$\begin{aligned} J_{M_{mn}} &= \langle \mathcal{G}_{M_{mn}} \rangle \quad \text{with} \\ \mathcal{G}_{M_{mn}} &= \{g(x)\} \cup \{f_{ijkl}(x) : 1 \leq i < k \leq m, 1 \leq j < l \leq n\} \end{aligned} \quad (9)$$

Recall that the convex hull of the set $\mathcal{R}$ in (8) forms the unit nuclear norm ball and by definition of the theta bodies,

$$\overline{\text{conv}(v_{\mathbb{R}}(J_{M_{mn}}))} \subseteq \cdots \subseteq TH_{k+1}(J_{M_{mn}}) \subseteq TH_k(J_{M_{mn}}) \subseteq \cdots \subseteq TH_1(J_{M_{mn}}).$$

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

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

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

Proof The set $\mathcal{G}_{M_{mn}}$ is clearly a basis for the ideal $J_{M_{mn}}$. By Proposition 1 in the [Appendix](#), we only need to check whether the S -polynomial (see Definition 11) satisfies $S(p, q) \rightarrow_{\mathcal{G}_{M_{mn}}} 0$ for all $p, q \in \mathcal{G}_{M_{mn}}$ whenever the leading monomials $LM(p)$ and $LM(q)$ 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 10).

Notice that $LM(g) = x_{11}^2$ and $LM(f_{ijkl}) = x_{il}x_{kj}$ 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 $S(f_{ijkl}, f_{i\hat{j}\hat{k}\hat{l}}) \rightarrow_{\mathcal{G}_{M_{mn}}} 0$ whenever the leading monomials $LM(f_{ijkl})$ and $LM(f_{i\hat{j}\hat{k}\hat{l}})$ are not relatively prime. First we consider

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

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

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

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

In order to show that $\mathcal{G}_{M_{mn}}$ is a reduced Gröbner basis (see Definition 9), we first notice that $LC(f) = 1$ for all $f \in \mathcal{G}_{M_{mn}}$. In addition, the leading monomial of $f \in \mathcal{G}_{M_{mn}}$ is always of degree two and there are no two different polynomials $f_i, f_j \in \mathcal{G}_{M_{mn}}$ such that $LM(f_i) = LM(f_j)$. Therefore, $\mathcal{G}_{M_{mn}}$ is the reduced Gröbner basis of the ideal $J_{M_{mn}}$ with respect to the grevlex order. $\square$

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

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

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

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

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

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

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

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

$$\min_{t \in \mathbb{R}, \mathbf{y} \in \mathbb{R}^{\mathcal{B}_k}} t \quad \text{subject to} \quad \mathbf{M}_{\mathcal{B}_k}[\mathbf{y}] \succcurlyeq 0, y_0 = t, \mathbf{y}_{\mathcal{B}_1} = \mathbf{X}. \quad (10)$$

The last constraint might require some explanation. The vector $\mathbf{y}_{\mathcal{B}_1}$ denotes the restriction of $\mathbf{y}$ to the indices in $\mathcal{B}_1$, where the latter can be identified with the set $[m] \times [n]$ indexing the matrix entries. Therefore, $\mathbf{y}_{\mathcal{B}_1} = \mathbf{X}$ means componentwise $y_{x_{11}+J} = X_{11}, y_{x_{12}+J} = X_{12}, \dots, y_{x_{mn}+J} = X_{mn}$. For the purpose of illustration, we focus on the θ_1 -norm in $\mathbb{R}^{2 \times 2}$ in Section 3.1 below, and provide a step-by-step procedure for building the corresponding semidefinite program in (10).

Notice that the number of elements in $\mathcal{B}_1$ is $mn + 1$, and in $\mathcal{B}_2 \setminus \mathcal{B}_1$ is $\frac{m \cdot (m+1)}{2} \cdot \frac{n \cdot (n+1)}{2} - 1 \sim \frac{(mn)^2}{2}$, i.e., the number of elements of the θ -basis restricted to the degree 2 scales polynomially in the total number of matrix entries mn . Therefore, the computational complexity of the SDP in (10) is polynomial in mn .

We will show next that the theta body $TH_1(J)$ 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 *Let $\mathcal{O}_c$ ($\mathcal{O}_r$) denote the set of all matrices $\mathbf{M} \in \mathbb{R}^{n \times m}$ with orthonormal columns (rows), i.e., $\mathcal{O}_c = \{\mathbf{M} \in \mathbb{R}^{n \times m} : \mathbf{M}^T \mathbf{M} = \mathbf{I}_m\}$ and $\mathcal{O}_r = \{\mathbf{M} \in \mathbb{R}^{n \times m} : \mathbf{M} \mathbf{M}^T = \mathbf{I}_n\}$. Then*

$$\{\mathbf{X} \in \mathbb{R}^{m \times n} : \|\mathbf{X}\|_* \leq 1\} = \{\mathbf{X} \in \mathbb{R}^{m \times n} : \text{tr}(\mathbf{M}\mathbf{X}) \leq 1, \text{ for all } \mathbf{M} \in \mathcal{O}_c \cup \mathcal{O}_r\}. \quad (11)$$

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

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

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

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

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

This completes the proof. $\square$

Next, using Lemma 3, we show that the theta body $TH_1(J)$ equals the nuclear norm ball. This result is related to Theorem 4.4 in [28].

Theorem 2 *The polynomial ideal $J_{M_{mn}}$ defined in (9) is TH_1 -exact, i.e.,*

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

In other words,

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

Proof By definition of $TH_1(J_{M_{mn}})$, it is enough to show that the boundary of the unit nuclear norm can be written as 1-sos mod $J_{M_{mn}}$, which by Lemma 3 means that the polynomial $1 - \sum_{i=1}^m \sum_{j=1}^n x_{ij} M_{ji}$ is 1-sos mod $J_{M_{mn}}$ for all $\mathbf{M} \in \mathcal{O}_c \cup \mathcal{O}_r$.

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

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

since

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

and

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

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

Next, we define transformed variables

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

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

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

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

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

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

The following corollary is a direct consequence of Theorem 2 and the nestedness property (5) of theta bodies.

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

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

Moreover,

$$TH_1(J_{M_{mn}}) = TH_2(J_{M_{mn}}) = \cdots = \text{conv}(v_{\mathbb{R}}(J_{M_{mn}})).$$

Remark 2 The ideal (9) is not the only choice that satisfies (6). The following polynomial ideal was suggested in [12],

$$J = \left\langle \{x_{ij} - u_i v_j\}_{i \in [m], j \in [n]}, \sum_{i=1}^m u_i^2 - 1, \sum_{j=1}^n v_j^2 - 1 \right\rangle \quad (12)$$

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

$$\begin{aligned} \mathcal{G} = & \left\{ g_1^{ij} = x_{ij} - u_i v_j : i \in [m], j \in [n] \right\} \cup \left\{ g_2 = \sum_{i=1}^m u_i^2 - 1, g_3 = \sum_{j=1}^n v_j^2 - 1 \right\} \\ & \cup \left\{ g_4^{i,j,k} = x_{ij}u_k - x_{kj}u_i : 1 \leq i < k \leq m, j \in [n] \right\} \\ & \cup \left\{ g_5^{i,j,k} = x_{ij}v_k - x_{ik}v_j : i \in [m], 1 \leq j < k \leq n \right\} \end{aligned}$$

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

Obviously, this Gröbner basis is much more complicated than the one of the ideal $J_{M_{mn}}$ introduced above. Therefore, computations (both theoretical and numerical) with this alternative ideal seem to be more demanding. In any case, the variables $\{u_i\}_{i=1}^m$ and $\{v_j\}_{j=1}^n$ are only auxiliary ones, so one would like to eliminate these from the above Gröbner basis. By doing so, one obtains the Gröbner basis $\mathcal{G}_{M_{mn}}$ defined in (9). Notice that $\sum_{i=1}^m \sum_{j=1}^n x_{ij}^2 - 1 = g_{13} + \sum_{i=2}^m g_{10}^i + \sum_{j=2}^n g_{11}^j$ together with $\{g_{12}^{i,j,k,l}\}$ form the basis $\mathcal{G}_{M_{mn}}$.

3.1 The θ_1 -norm in $\mathbb{R}^{2 \times 2}$

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

$$J = J_{M_{22}} = \left\langle x_{12}x_{21} - x_{11}x_{22}, x_{11}^2 + x_{12}^2 + x_{21}^2 + x_{22}^2 - 1 \right\rangle \tag{14}$$

The associated algebraic variety is of the form

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

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

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

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

$$\begin{aligned}\mathcal{B}_1 &= \{1 + J, x_{11} + J, x_{12} + J, x_{21} + J, x_{22} + J\} \\ \mathcal{B}_2 &= \mathcal{B}_1 \cup \{x_{11}x_{12} + J, x_{11}x_{21} + J, x_{11}x_{22} + J, x_{12}^2 + J, x_{12}x_{22} + J, \\ &\quad x_{21}^2 + J, x_{21}x_{22} + J, x_{22}^2 + J\}.\end{aligned}$$

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

Linearizing the elements of $\mathcal{B}_2$ results in Table 1, where the monomials f in the first row stand for an element $f + J \in \mathcal{B}_2$.

Therefore, $[\mathbf{x}]_{\mathcal{B}_1} = (1, x_{11}, x_{12}, x_{21}, x_{22})^T$ and the following combinatorial moment matrix $\mathbf{M}_{\mathcal{B}_1}(\mathbf{x}, \mathbf{y})$ (see Definition 4) is given as

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

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

By Theorem 1, the first theta body $TH_1(J)$ is the closure of

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

where $\pi_{\mathbf{x}}$ represents the projection onto the variables $\mathbf{x}$, i.e., the projection onto $x_{11}, x_{12}, x_{21}, x_{22}$. Furthermore, θ_1 -norm of a matrix $\mathbf{X} \in \mathbb{R}^{2 \times 2}$ induced by the $TH_1(J)$ and denoted as $\|\cdot\|_{\theta_1}$ can be computed as

$$\|\mathbf{X}\|_{\theta_1} = \inf t \text{ s.t. } \mathbf{X} \in t\mathbf{Q}_{\mathcal{B}_1}(J) \quad (15)$$

Table 1 Linearization of the elements of $\mathcal{B}_2 = \{f + J\}$ for matrix 2×2 case

$1 + J$	$x_{11} + J$	$x_{12} + J$	$x_{21} + J$	$x_{22} + J$	$x_{11}x_{12} + J$	$x_{11}x_{21} + J$
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	

which is equivalent to

$$\inf_{t \in \mathbb{R}, \mathbf{y} \in \mathbb{R}^8} t \quad \text{s.t.} \quad \mathbf{M} = \begin{bmatrix} t & X_{11} & X_{12} & X_{21} & X_{22} \\ X_{11} & -y_4 - y_6 - y_8 + t & y_1 & y_2 & y_3 \\ X_{12} & y_1 & y_4 & y_3 & y_5 \\ X_{21} & y_2 & y_3 & y_6 & y_7 \\ X_{22} & y_3 & y_5 & y_7 & y_8 \end{bmatrix} \succeq 0. \quad (16)$$

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

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

Remark 3 In compressive sensing, reconstruction of sparse signals via ℓ_1 -norm minimization is well-understood (see, for example, [10, 20, 23]). It is possible to provide hierarchical relaxations via theta bodies of the unit ℓ_1 -norm ball. However, as in the matrix scenario discussed above, all these relaxations coincide with the unit ℓ_1 -norm ball, [58].

4 The tensor θ_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 θ_k -norms are equal to the matrix nuclear norm, their generalization to the tensor case may all be viewed as natural generalizations of the nuclear norm. We focus mostly on the θ_1 -norm whose unit norm ball is the largest in a hierarchical sequence of relaxations. Unlike in the matrix case, the θ_1 -norm defines a new tensor norm, that up to the best of our knowledge has not been studied before.

The polynomial ideal will be generated by the minors of order two of the unfoldings—and matricizations in the case $d \geq 4$ —of the tensors, where each variable corresponds to one entry in the tensor. As we will see, a tensor is of rank one if and only if all order-two minors of the unfoldings (matricizations) vanish. While the order-three case requires to consider all three unfoldings, there are several possibilities for the order- d case when $d \geq 4$. In fact, a d th-order tensor is of rank one if all minors of all unfoldings vanish so that it may be enough to consider only the unfoldings. However, one may as well consider the ideal generated by all minors of *all matricizations* or one may consider a subset of matricizations including all unfoldings. Indeed, any *tensor format*—and thereby any notion of tensor rank—corresponds to a set of matricizations and in this way, one may associate a θ_k -norm to a certain tensor format. We refer to, e.g., [33, 53] 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 will mainly concentrate on the case that *all* matricizations are taken into account for

defining the ideal. Only for the case $d = 4$, we will briefly discuss the case, that the ideal is generated only by the minors corresponding to the four unfoldings.

Below, we consider first the special case of third-order tensors and continue then with fourth-order tensors. In Section 4.2 we will treat the general d th-order case.

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

$$\begin{aligned}\mathcal{S}_1 &= \{(\alpha, \beta) : 1 \leq \alpha_1 < \beta_1 \leq n_1, 1 \leq \beta_2 < \alpha_2 \leq n_2, 1 \leq \beta_3 \leq \alpha_3 \leq n_3\}, \\ \mathcal{S}_2 &= \{(\alpha, \beta) : 1 \leq \alpha_1 \leq \beta_1 \leq n_1, 1 \leq \beta_2 < \alpha_2 \leq n_2, 1 \leq \alpha_3 < \beta_3 \leq n_3\}, \\ \mathcal{S}_3 &= \{(\alpha, \beta) : 1 \leq \alpha_1 < \beta_1 \leq n_1, 1 \leq \alpha_2 \leq \beta_2 \leq n_2, 1 \leq \beta_3 < \alpha_3 \leq n_3\}, \\ \overline{\mathcal{S}}_i &= \{(\alpha, \beta) : (\alpha, \beta) \in \mathcal{S}_i \text{ and } \alpha_j \neq \beta_j, \text{ for all } j \in [3]\}, \quad \text{for all } i \in [3].\end{aligned}$$

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

$$\begin{aligned}f^{(\alpha, \beta)}(\mathbf{x}) &= x_\alpha x_\beta - x_{\alpha \vee \beta} x_{\alpha \wedge \beta}, \quad (\alpha, \beta) \in \mathcal{S} := \mathcal{S}_1 \cup \mathcal{S}_2 \cup \mathcal{S}_3 \\ g_3(\mathbf{x}) &= \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} x_{ijk}^2 - 1,\end{aligned}$$

where $[\alpha \vee \beta]_i = \max\{\alpha_i, \beta_i\}$ and $[\alpha \wedge \beta]_i = \min\{\alpha_i, \beta_i\}$. In particular, the following order-two minor of $\mathbf{X}^{(1)}$ is not contained in $\{f^{(\alpha, \beta)} : (\alpha, \beta) \in \mathcal{S}\}$

$$f = x_\alpha x_\beta - x_{\hat{\alpha}} x_{\hat{\beta}}, \quad \text{where } \hat{\alpha} = (\alpha_1, \beta_2, \beta_3), \hat{\beta} = (\beta_1, \alpha_2, \alpha_3) \text{ and } (\alpha, \beta) \in \overline{\mathcal{S}}_3.$$

We remark that in real algebraic geometry and commutative algebra, polynomials $f^{(\alpha, \beta)}$ are known as Hibi relations (see [34]).

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

$$g_3(\mathbf{X}) = 0 \text{ and } f^{(\alpha, \beta)}(\mathbf{X}) = 0 \text{ for all } (\alpha, \beta) \in \mathcal{S}. \quad (17)$$

Proof Sufficiency of (17) 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 fibers $\mathbf{X}_{\cdot \alpha_2 \alpha_3}$ and $\mathbf{X}_{\cdot \beta_2 \beta_3}$ satisfy

$$\begin{aligned}-X_{n_1 \alpha_2 \alpha_3} \begin{bmatrix} X_{1 \beta_2 \beta_3} \\ X_{2 \beta_2 \beta_3} \\ \vdots \\ X_{n_1 \beta_2 \beta_3} \end{bmatrix} + X_{n_1 \beta_2 \beta_3} \begin{bmatrix} X_{1 \alpha_2 \alpha_3} \\ X_{2 \alpha_2 \alpha_3} \\ \vdots \\ X_{n_1 \alpha_2 \alpha_3} \end{bmatrix} \\ = \begin{bmatrix} -X_{1 \beta_2 \beta_3} X_{n_1 \alpha_2 \alpha_3} + X_{1 \beta_2 \beta_3} X_{n_1 \alpha_2 \alpha_3} \\ -X_{2 \beta_2 \beta_3} X_{n_1 \alpha_2 \alpha_3} + X_{2 \beta_2 \beta_3} X_{n_1 \alpha_2 \alpha_3} \\ \vdots \\ -X_{n_1 \beta_2 \beta_3} X_{n_1 \alpha_2 \alpha_3} + X_{n_1 \beta_2 \beta_3} X_{n_1 \alpha_2 \alpha_3} \end{bmatrix} = \mathbf{0},\end{aligned}$$

where we used that $f^{(\alpha, \beta)}(\mathbf{X}) = 0$ for all $(\alpha, \beta) \in \mathcal{S}$. From $g_3(\mathbf{X}) = 0$ it follows that the tensor $\mathbf{X}$ is normalized.

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

A third-order tensor $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is rank one if and only if all three unfoldings $\mathbf{X}^{(1)} \in \mathbb{R}^{n_1 \times n_2 n_3}$, $\mathbf{X}^{(2)} \in \mathbb{R}^{n_2 \times n_1 n_3}$, and $\mathbf{X}^{(3)} \in \mathbb{R}^{n_3 \times n_1 n_2}$ are rank-one matrices. Notice that $f^{(\alpha, \beta)}(\mathbf{X}) = 0$ for all $(\alpha, \beta) \in \mathcal{S}$ is equivalent to the statement that the ℓ -th unfolding $\mathbf{X}^{(\ell)}$ 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 $J_3 \subset \mathbb{R}[\mathbf{x}] = \mathbb{R}[x_{111}, x_{112}, \dots, x_{n_1 n_2 n_3}]$ as the one generated by

$$\mathcal{G}_3 = \left\{ f^{(\alpha, \beta)}(\mathbf{x}) : (\alpha, \beta) \in \mathcal{S} \right\} \cup \{g_3(\mathbf{x})\}, \quad (18)$$

i.e., $J_3 = \langle \mathcal{G}_3 \rangle$. 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 3 *The basis $\mathcal{G}_3$ defined in (18) forms the reduced Gröbner basis of the ideal $J_3 = \langle \mathcal{G}_3 \rangle$ with respect to the grevlex order.*

Proof Similarly to the proof of Theorem 2 we need to show that $S(p, q) \rightarrow_{\mathcal{G}_3} 0$ for all polynomials $p, q \in \mathcal{G}_3$ whose leading terms are not relatively prime. The leading monomials with respect to the grevlex ordering are given by

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

The leading terms of g_3 and $f^{(\alpha, \beta)}$ are always relatively prime. First we consider two distinct polynomials $f, g \in \{f^{(\alpha, \beta)} : (\alpha, \beta) \in \mathcal{S}_3\}$. Let $f = f^{(\alpha, \beta)}$ and $g = f^{(\alpha, \bar{\beta})}$ for $(\alpha, \beta) \in \mathcal{S}_3$, where $\bar{\beta} = (\beta_1, \alpha_2, \beta_3)$. That is,

$$f(\mathbf{x}) = x_\alpha x_\beta - x_{\alpha \vee \beta} x_{\alpha \wedge \beta}, \quad g(\mathbf{x}) = x_\alpha x_{\bar{\beta}} - x_{\alpha \vee \bar{\beta}} x_{\alpha \wedge \bar{\beta}}.$$

Since $\alpha \wedge \beta = \alpha \wedge \bar{\beta}$ and $f^{(\beta, \alpha \vee \bar{\beta})} \in \{f^{(\alpha, \beta)} : (\alpha, \beta) \in \mathcal{S}_2\}$, then

$$S(f, g) = x_{\alpha \wedge \beta} \left(-x_{\bar{\beta}} x_{\alpha \vee \beta} + x_\beta x_{\alpha \vee \bar{\beta}} \right) = x_{\alpha \wedge \beta} f^{(\beta, \alpha \vee \bar{\beta})} \rightarrow_{\mathcal{G}_3} 0.$$

Next we show that $S(f, g) \in J_3$, for $f \in \{f^{(\alpha, \beta)} : (\alpha, \beta) \in \mathcal{S}_2\}$ and $g \in \{f^{(\alpha, \beta)} : (\alpha, \beta) \in \mathcal{S}_1\}$. Let $f = f^{(\alpha, \hat{\beta})}$ with $\hat{\beta} = (\alpha_1, \beta_2, \beta_3)$ and $g = f^{(\alpha, \tilde{\beta})}$ with $\tilde{\beta} = (\beta_1, \beta_2, \alpha_3)$, where $(\alpha, \beta) \in \mathcal{S}_2$. Since $x_{\alpha \wedge \hat{\beta}} = x_{\alpha \wedge \tilde{\beta}}$, $f^{(\hat{\beta}, \alpha \vee \tilde{\beta})} \in \{f^{(\alpha, \beta)} : (\alpha, \beta) \in \mathcal{S}_3\}$, and $f^{(\alpha \vee \hat{\beta}, \tilde{\beta})} \in \{f^{(\alpha, \beta)} : (\alpha, \beta) \in \mathcal{S}_1\}$

$$S(f, g) = x_{\alpha \wedge \hat{\beta}} \left(-x_{\tilde{\beta}} x_{\alpha \vee \hat{\beta}} + x_{\hat{\beta}} x_{\alpha \vee \tilde{\beta}} \right) = x_{\alpha \wedge \hat{\beta}} \left(f^{(\hat{\beta}, \alpha \vee \tilde{\beta})} - f^{(\alpha \vee \hat{\beta}, \tilde{\beta})} \right) \rightarrow_{\mathcal{G}_3} 0.$$

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

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 $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ (not considering the same minor twice). One might think that the θ_1 -norm obtained in this way corresponds to a (weighted) sum of the nuclear norms of the unfoldings, which has been used in [25, 39] for tensor recovery. The examples of cubic tensors $\mathbf{X} \in \mathbb{R}^{2 \times 2 \times 2}$ presented in Table 2 show that this is not the case. Assuming that θ_1 -norm is a linear combination of the nuclear norm of the unfoldings, there exist $\alpha, \beta, \gamma \in \mathbb{R}$ such that $\alpha \|\mathbf{X}^{(1)}\|_* + \beta \|\mathbf{X}^{(2)}\|_* + \gamma \|\mathbf{X}^{(3)}\|_* = \|\mathbf{X}\|_{\theta_1}$. From the first and the second tensors in Table 2 we obtain $\gamma = 0$. Similarly, the first and the third tensors and the first and the fourth tensors give $\beta = 0$ and $\alpha = 0$, respectively. Thus, the θ_1 -norm does not coincide with a weighted sum of the nuclear norms of the unfoldings. In addition, the last tensor shows that the θ_1 -norm does not equal maximum of the norms of the unfoldings.

Theorem 3 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 generated by the following polynomials

$$\begin{aligned} f_{(\alpha, \beta)}^{(1)}(\mathbf{x}) &= -x_{\alpha_1 \alpha_2 \alpha_3} x_{\beta_1 \beta_2 \beta_3} + x_{\alpha_1 \beta_2 \beta_3} x_{\beta_1 \alpha_2 \alpha_3}, \quad \text{for } (\alpha, \beta) \in \mathcal{T}^{(1)} \\ f_{(\alpha, \beta)}^{(2)}(\mathbf{x}) &= -x_{\alpha_1 \alpha_2 \alpha_3} x_{\beta_1 \beta_2 \beta_3} + x_{\beta_1 \alpha_2 \beta_3} x_{\alpha_1 \beta_2 \alpha_3}, \quad \text{for } (\alpha, \beta) \in \mathcal{T}^{(2)} \\ f_{(\alpha, \beta)}^{(3)}(\mathbf{x}) &= -x_{\alpha_1 \alpha_2 \alpha_3} x_{\beta_1 \beta_2 \beta_3} + x_{\beta_1 \beta_2 \alpha_3} x_{\alpha_1 \alpha_2 \beta_3}, \quad \text{for } (\alpha, \beta) \in \mathcal{T}^{(3)}, \end{aligned}$$

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

	$\mathbf{X} \in \mathbb{R}^{2 \times 2 \times 2}$	$\ \mathbf{X}^{(1)}\ _*$	$\ \mathbf{X}^{(2)}\ _*$	$\ \mathbf{X}^{(3)}\ _*$	$\ \mathbf{X}\ _{\theta_1}$
1	$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$	2	2	2	2
2	$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$	2	2	$\sqrt{2}$	2
3	$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$	2	$\sqrt{2}$	2	2
4	$\begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$	$\sqrt{2}$	2	2	2
5	$\begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}$	$\sqrt{2} + 1$	$\sqrt{2} + 1$	$\sqrt{2} + 1$	3

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

$$\mathcal{T}^{[k]} = \{(\alpha, \beta) : \alpha_k \neq \beta_k, \bar{\alpha} \neq \bar{\beta}, \text{ where } \bar{\alpha}_k = \bar{\beta}_k = 0, \bar{\alpha}_\ell = \alpha_\ell, \bar{\beta}_\ell = \beta_\ell\}.$$

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

$$\begin{aligned} f_{(\alpha,\beta)}^{[1]}(\mathbf{x}) &= -x_{\alpha_1\alpha_2\alpha_3}x_{\beta_1\beta_2\beta_3} + x_{\alpha_1\beta_2\beta_3}x_{\beta_1\alpha_2\alpha_3}, \quad \text{for } (\alpha, \beta) \in \mathcal{T}^{[1]}, \\ f_{(\alpha,\beta)}^{[1,2]}(\mathbf{x}) &= -x_{\alpha_1\alpha_2\alpha_3}x_{\beta_1\beta_2\beta_3} + x_{\alpha_1\alpha_2\beta_3}x_{\beta_1\beta_2\alpha_3}, \quad \text{for } (\alpha, \beta) \in \mathcal{T}^{[1,2]}, \end{aligned}$$

where $\mathcal{T}^{[1,2]} = \{(\alpha, \beta) : (\alpha_1, \alpha_2, 0) \neq (\beta_1, \beta_2, 0), \alpha_3 \neq \beta_3\}$.

Theorem 4 *The polynomial ideals J_3 and $J_{3,\text{TT}}$ are equal.*

Remark 5 As a consequence, $\mathcal{G}_3$ is also the reduced Gröbner basis for the ideal $J_{3,\text{TT}}$ with respect to the grevlex ordering.

Proof Notice that $(\mathbf{X}^{[3]})^T = \mathbf{X}^{[1,2]}$ and therefore

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

Hence, it is enough to show that $f_{(\alpha,\beta)}^{[2]} \in J_{3,\text{TT}}$, for all $(\alpha, \beta) \in \mathcal{T}^{[2]}$. By definition of $\mathcal{T}^{[2]}$, we have that $\alpha_2 \neq \beta_2$ and $(\alpha_1, 0, \alpha_3) \neq (\beta_1, 0, \beta_3)$. 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

$$\begin{aligned} f(\mathbf{x}) &= -x_{\alpha_1\alpha_2\alpha_3}x_{\beta_1\beta_2\beta_3} + x_{\beta_1\alpha_2\beta_3}x_{\alpha_1\beta_2\alpha_3}, \quad (\alpha, \beta) \in \mathcal{T}^{[2]} \\ g(\mathbf{x}) &= -x_{\beta_1\beta_2\alpha_3}x_{\alpha_1\alpha_2\beta_3} + x_{\beta_1\alpha_2\beta_3}x_{\alpha_1\beta_2\alpha_3}, \quad (\beta_1, \beta_2, \alpha_3, \alpha_1, \alpha_2, \beta_3) \in \mathcal{T}^{[1]} \\ h(\mathbf{x}) &= -x_{\alpha_1\alpha_2\alpha_3}x_{\beta_1\beta_2\beta_3} + x_{\alpha_1\alpha_2\beta_3}x_{\beta_1\beta_2\alpha_3}, \quad (\alpha, \beta) \in \mathcal{T}^{[1,2]}. \end{aligned}$$

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

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 $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ is of rank-one if and only if all its matricizations are rank-one matrices, or equivalently, if all minors of order two of each matricization vanish.

The description of the polynomial ideal generated by the second-order minors of all matricizations of a tensor $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times \dots \times n_d}$ 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 $(\alpha_1, \alpha_2, \dots, \alpha_d)$ of a tensor $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times \dots \times n_d}$ corresponds to the variable $x_{\alpha_1 \alpha_2 \dots \alpha_d}$ or simply x_α . We aim at introducing a set of polynomials of the form

$$f_d^{(\alpha, \beta)}(\mathbf{x}) := -x_{\alpha \wedge \beta} x_{\alpha \vee \beta} + x_\alpha x_\beta \quad (19)$$

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

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

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 α and β for forming the minors in (19). This implies that, without loss of generality,

$$\begin{aligned} \alpha_j &> \beta_j, & \text{for all } j \in \mathcal{M} \\ \alpha_k &< \beta_k, & \text{for all } k \in \mathcal{S} \setminus \mathcal{M}. \end{aligned}$$

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 $\mathbf{X}^{\mathcal{M}}$ coincides with the set of all two-minors of $\mathbf{X}^{\mathcal{S} \setminus \mathcal{M}}$.

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

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

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

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

$$\begin{aligned} \mathcal{I}_d^{\mathcal{S}, \mathcal{M}} &:= \{(\alpha, \beta) : \alpha_i = \beta_i, \text{ for all } i \notin \mathcal{S} \\ &\quad \alpha_j > \beta_j, \text{ for all } j \in \mathcal{M} \\ &\quad \alpha_k < \beta_k, \text{ for all } k \in \mathcal{S} \setminus \mathcal{M}\}. \end{aligned}$$

For notational purposes, we define

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

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

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

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

$$\mathcal{G}_d = \bigcup_{\mathcal{S} \in \mathcal{S}_{[d]}} \{f_d^{\mathcal{S}}\} \cup \{g_d\} \subset \mathbb{R}[\mathbf{x}] = \mathbb{R}[x_{11\dots 1}, x_{11\dots 2}, \dots, x_{n_1 n_2 \dots n_d}],$$

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

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

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

Proof Again, we use Buchberger’s criterion stated in Theorem 9. First notice that the polynomials g_d and $f_d^{(\alpha, \beta)}$ are always relatively prime, since $LM(g_d) = x_{11\dots 1}^2$ and $LM(f_d^{(\alpha, \beta)}) = x_{\alpha}x_{\beta}$ for $(\alpha, \beta) \in \mathcal{T}_d^{\mathcal{M}, \mathcal{S}}$, where $\mathcal{S} \in \mathcal{S}_{[d]}$ and $\mathcal{M} \in \mathcal{P}_{\mathcal{S}}$. 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 $LM(f_1) \neq LM(f_2)$. If these leading monomials are not relatively prime, the S -polynomial is of the form

$$S(f_1, f_2) = x_{\alpha^1}x_{\alpha^2}x_{\alpha^3} - x_{\bar{\alpha}^1}x_{\bar{\alpha}^2}x_{\bar{\alpha}^3}$$

with $\{\alpha_k^1, \alpha_k^2, \alpha_k^3\} = \{\bar{\alpha}_k^1, \bar{\alpha}_k^2, \bar{\alpha}_k^3\}$ 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 ℓ 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 $\hat{\alpha}_k^{1,i} \leq \hat{\alpha}_k^{2,i} \leq \hat{\alpha}_k^{3,i}$, 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 $f_1, f_2 \in \mathcal{G}_d$

$$S^0 = S(f_1, f_2) = x_{\alpha^1} x_{\alpha^2} x_{\alpha^3} - x_{\hat{\alpha}^1} x_{\hat{\alpha}^2} x_{\hat{\alpha}^3}, i = 0$$

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

1) **Let** $LM(S^i) = x_{\hat{\alpha}^1, i} x_{\hat{\alpha}^2, i} x_{\hat{\alpha}^3, i}$ **and** $\mathbb{N}LM(S^i) = |S^i - LT(S^i)|$

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

$$\alpha_k^{1, i} < \alpha_k^{2, i} \quad \text{and} \quad \alpha_\ell^{1, i} > \alpha_\ell^{2, i} \quad \text{s.t.} \quad \mathcal{M}_i := \left\{ \ell \in [d] : \alpha_\ell^{1, i} > \alpha_\ell^{2, i} \right\} \in \mathcal{P}_{\mathcal{S}},$$

where $\mathcal{S} := \left\{ k \in [d] : \alpha_k^{1, i} \neq \alpha_k^{2, i} \right\}$ and let $\alpha^{3, i}$ be the remaining index in $\{\hat{\alpha}^1, \hat{\alpha}^2, \hat{\alpha}^3\} \setminus \{\alpha^{1, i}, \alpha^{2, i}\}$.

3) Divide S^i by $f_d^{(\alpha^{1, i}, \alpha^{2, i})} = x_{\alpha^{1, i}} x_{\alpha^{2, i}} - x_{\alpha^{1, i} \wedge \alpha^{2, i}} x_{\alpha^{1, i} \vee \alpha^{2, i}}$ to obtain

$$S^i = LC(S^i) \left[x_{\alpha^{3, i}} (-x_{\alpha^{1, i} \wedge \alpha^{2, i}} x_{\alpha^{1, i} \vee \alpha^{2, i}} + x_{\alpha^{1, i}} x_{\alpha^{2, i}}) \right. \\ \left. + x_{\alpha^{1, i} \wedge \alpha^{2, i}} x_{\alpha^{1, i} \vee \alpha^{2, i}} x_{\alpha^{3, i}} - \mathbb{N}LM(S^i) \right].$$

4) Define

$$S^{i+1} := x_{\alpha^{1, i} \wedge \alpha^{2, i}} x_{\alpha^{1, i} \vee \alpha^{2, i}} x_{\alpha^{3, i}} - \mathbb{N}LM(S^i).$$

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 $f_1(\mathbf{x}) := f_4^{(1212, 2123)}(\mathbf{x}) = -x_{1112}x_{2223} + x_{1212}x_{2123} \in \mathcal{G}_4$ (with the corresponding sets $\mathcal{S} = \{1, 2, 3, 4\}$, $\mathcal{M} = \{2\}$) and $f_2(\mathbf{x}) := f_4^{(3311, 2123)}(\mathbf{x}) = -x_{2111}x_{3323} + x_{3311}x_{2123} \in \mathcal{G}_4$ (with the corresponding sets $\mathcal{S} = \{1, 2, 3, 4\}$, $\mathcal{M} = \{1, 2\}$). We will show that $S(f_1, f_2) = -x_{1112}x_{2223}x_{3311} + x_{1212}x_{2111}x_{3323} \rightarrow_{\mathcal{G}_4} 0$ by going through the division algorithm.

In iteration $i = 0$ we set $S^0 = S(f_1, f_2) = -x_{1112}x_{2223}x_{3311} + x_{1212}x_{2111}x_{3323}$. The leading monomial is $LM(S^0) = x_{1112}x_{2223}x_{3311}$, the leading coefficient is $LC(S^0) = -1$, and the non-leading monomial is $\mathbb{N}LM(S^0) = x_{1212}x_{2111}x_{3323}$. 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 $x_{\alpha^{1, 0}} x_{\alpha^{2, 0}} - x_{\alpha^{1, 0} \wedge \alpha^{2, 0}} x_{\alpha^{1, 0} \vee \alpha^{2, 0}}$ then equals the polynomial $f_4^{(1112, 3311)}(\mathbf{x}) = -x_{1111}x_{3312} + x_{1112}x_{3311} \in \mathcal{G}_4$ and we can write

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

The leading and non-leading monomials of S^1 are $LM(S^1) = x_{1111}x_{2223}x_{3312}$ and $\mathbb{N}LM(S^1) = x_{1212}x_{2111}x_{3323}$, respectively, while $LC(S^1) = 1$. 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 $x_{\alpha^{1,1}}x_{\alpha^{2,1}} - x_{\alpha^{1,1} \wedge \alpha^{2,1}}x_{\alpha^{1,1} \vee \alpha^{2,1}}$ in the step 4) equals $f_4^{(3312,2223)}(\mathbf{x}) = -x_{2212}x_{3323} + x_{3312}x_{2223} \in \mathcal{G}_4$ and we obtain

$$S^1 = 1 \cdot \left(x_{1111}(-x_{2212}x_{3323} + x_{2223}x_{3312}) + \underbrace{x_{1111}x_{2212}x_{3323} - x_{1212}x_{2111}x_{3323}}_{= S^2} \right).$$

The index sets of the monomial $x_{\alpha^1}x_{\alpha^2}x_{\alpha^3} = x_{1111}x_{2212}x_{3323}$ in S^2 satisfy

$$\alpha_k^1 \leq \alpha_k^2 \leq \alpha_k^3 \quad \text{for all } k \in [4]$$

and therefore it is the non-leading monomial of S^2 , i.e., $\mathbb{N}LM(S^2) = x_{1111}x_{2212}x_{3323}$. Thus, $LM(S^2) = x_{1212}x_{2111}x_{3323}$ and $LC(S^2(f_1, f_2)) = -1$. 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

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

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

$$S(f_1, f_2) = S^0 = LC(S^0)x_{2223}f_4^{(1112,3311)}(\mathbf{x}) + LC(S^0)LC(S^1)x_{1111}f_4^{(3312,2223)}(\mathbf{x}) + LC(S^0)LC(S^1)LC(S^2)x_{3323}f_4^{(2111,1212)}(\mathbf{x}).$$

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

Let us now return to the general proof. We first show that there always exist indices $\alpha^{1,i}, \alpha^{2,i}$ satisfying line 7 of Algorithm 2 unless $S^i = 0$. We start by setting $\mathbf{x}^{\alpha^i} = x_{\hat{\alpha}^{1,i}}x_{\hat{\alpha}^{2,i}}x_{\hat{\alpha}^{3,i}}$ with $x_{\hat{\alpha}^{1,i}} \geq x_{\hat{\alpha}^{2,i}} \geq x_{\hat{\alpha}^{3,i}}$ to be the leading monomial and $\mathbf{x}^{\beta^i}$ to be the non-leading monomial of S^i . The existence of a polynomial $h \in \mathcal{G}_d$ such that $LM(h)$ divides $LM(S^i) = x_{\hat{\alpha}^{1,i}}x_{\hat{\alpha}^{2,i}}x_{\hat{\alpha}^{3,i}} = \mathbf{x}^{\alpha^i}$ is equivalent to the existence of $\alpha^{1,i}, \alpha^{2,i} \in \{\hat{\alpha}^{1,i}, \hat{\alpha}^{2,i}, \hat{\alpha}^{3,i}\}$ such that there exists at least one k and at least one ℓ for which $\alpha_k^{1,i} < \alpha_k^{2,i}$ and $\alpha_\ell^{1,i} > \alpha_\ell^{2,i}$. If such pair does not exist in iteration i , we have

$$\hat{\alpha}_k^{1,i} \leq \hat{\alpha}_k^{2,i} \leq \hat{\alpha}_k^{3,i} \quad \text{for all } k \in [d]. \quad (20)$$

We claim that this cannot happen if $S^i \neq 0$. In fact, (20) would imply that the monomial $\mathbf{x}^{\alpha^i} = x_{\hat{\alpha}^{1,i}}x_{\hat{\alpha}^{2,i}}x_{\hat{\alpha}^{3,i}}$ is the smallest monomial $x_\beta x_\gamma x_\eta$ (with respect to the grevlex order) which satisfies

$$\{\beta_k, \gamma_k, \eta_k\} = \{\hat{\alpha}_k^{1,i}, \hat{\alpha}_k^{2,i}, \hat{\alpha}_k^{3,i}\} \quad \text{for all } k \in [d].$$

However, then $\mathbf{x}^{\alpha^i}$ would not be the leading monomial by definition of the grevlex order, which leads to a contradiction. Hence, we can always find indices $\alpha^{1,i}, \alpha^{2,i}$ satisfying line 7 in step 2) of Algorithm 2 unless $S^i = 0$.

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

$$h(\mathbf{x}) = f_d^{(\alpha^{1,0}, \alpha^{2,0})} = x_{\alpha^{1,0}}x_{\alpha^{2,0}} - x_{\alpha^{1,0} \wedge \alpha^{2,0}}x_{\alpha^{1,0} \vee \alpha^{2,0}} \in \mathcal{G}_d.$$

The division of S^0 by h results in

$$S^0 = LC(S^0) \left(x_{\alpha^{3,0}} \cdot f_d^{(\alpha^{1,0}, \alpha^{2,0})} + \underbrace{x_{\alpha^{1,0} \wedge \alpha^{2,0}} x_{\alpha^{1,0} \vee \alpha^{2,0}} x_{\alpha^{3,0}} - \text{NLM}(S^0)}_{= S^1} \right).$$

Note that by construction

$$\left[\alpha^{1,0} \wedge \alpha^{2,0} \right]_k \leq \left[\alpha^{1,0} \vee \alpha^{2,0} \right]_k \quad \text{for all } k \in [d]. \quad (21)$$

If $S^1 \neq 0$, then in the following iteration $i = 1$ we can assume $LM(S^1) = x_{\alpha^{1,0} \wedge \alpha^{2,0}} x_{\alpha^{1,0} \wedge \alpha^{2,0}} x_{\alpha^{3,0}}$. Due to (21), 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

$$S^1 = LC(S^1) \left(x_{\alpha^{3,1}} \cdot f_d^{(\alpha^{1,1}, \alpha^{2,1})} + \underbrace{x_{\alpha^{1,1} \wedge \alpha^{2,1}} x_{\alpha^{1,1} \vee \alpha^{2,1}} x_{\alpha^{3,1}} - \text{NLM}(S^0)}_{= S^2} \right)$$

with

$$\left[\alpha^{1,1} \wedge \alpha^{2,1} \right]_k \leq \left[\alpha^{3,1} \right]_k, \left[\alpha^{1,1} \vee \alpha^{2,1} \right]_k \quad \text{for all } k \in [d], \text{ and } x_{\alpha^{3,1}} = x_{\alpha^{1,0} \vee \alpha^{2,0}}.$$

Next, if $S^2 \neq 0$ and $LM(S^2) = x_{\alpha^{1,1} \wedge \alpha^{2,1}} x_{\alpha^{1,1} \vee \alpha^{2,1}} x_{\alpha^{3,1}}$ 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

$$S^2 = LC(S^2) \left(x_{\alpha^{3,2}} \cdot f_d^{(\alpha^{1,2}, \alpha^{2,2})} + \underbrace{x_{\alpha^{1,2} \wedge \alpha^{2,2}} x_{\alpha^{1,2} \vee \alpha^{2,2}} x_{\alpha^{3,2}} - \text{NLM}(S^0)}_{= S^3} \right)$$

such that

$$\left[\alpha^{3,2} \right]_k \leq \left[\alpha^{1,2} \wedge \alpha^{2,2} \right]_k \leq \left[\alpha^{1,2} \vee \alpha^{2,2} \right]_k \quad \text{for all } k \in [d], \text{ and } x_{\alpha^{3,2}} = x_{\alpha^{1,1} \wedge \alpha^{2,1}}.$$

Thus, in iteration $i = 3$ the leading monomial $LM(S^3)$ 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

$$S(f_1, f_2) = \sum_{i=0}^p \left(\prod_{j=0}^i LC(S^j) \right) x_{\alpha^{3,i}} \cdot f_d^{(\alpha^{1,i}, \alpha^{2,i})},$$

where $f_d^{(\alpha^{1,i}, \alpha^{2,i})} = -x_{\alpha^{1,i} \wedge \alpha^{2,i}} x_{\alpha^{1,i} \vee \alpha^{2,i}} + x_{\alpha^{1,i}} x_{\alpha^{2,i}} \in \mathcal{G}_d$ 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 $LC(g) = 1$ 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 $f_d^{(\alpha^1, \beta^1)}, f_d^{(\alpha^2, \beta^2)} \in \mathcal{G}_d$ (related to $\mathcal{S}_1, \mathcal{M}_1$ and $\mathcal{S}_2, \mathcal{M}_2$, respectively) it holds that $LM(f_d^{(\alpha^1, \beta^1)}) \neq LM(f_d^{(\alpha^2, \beta^2)})$ with $(\alpha^k, \beta^k) \in \mathcal{T}_d^{\mathcal{S}_k, \mathcal{M}_k}$ for $k = 1, 2$. But this

follows from the fact that all elements of $\mathcal{G}_d$ are different as remarked before the statement of the theorem. $\square$

We define the tensor θ_k -norm analogously to the matrix scenario.

Definition 6 The tensor θ_k -norm, denoted by $\|\cdot\|_{\theta_k}$, is the norm induced by the k -theta body $TH_k(J_d)$, i.e.,

$$\|\mathbf{X}\|_{\theta_k} = \inf \{t : \mathbf{X} \in tTH_k(J_d)\}.$$

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

$$\|\mathbf{X}\|_{\theta_k} = \min t \quad \text{subject to } \mathbf{X} \in t\mathbf{Q}_{\mathcal{B}_k}(J_d).$$

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

$$\min_{t \in \mathbb{R}, \mathbf{y} \in \mathbb{R}^{\mathcal{B}_k}} t \quad \text{subject to} \quad \mathbf{M}_{\mathcal{B}_k}[\mathbf{y}] \succcurlyeq 0, y_0 = t, \mathbf{y}_{\mathcal{B}_1} = \mathbf{X}. \quad (22)$$

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 $\mathbf{r}_{HOSVD} = (r_1, r_2, \dots, r_d)$ such that $r_i = \text{rank}(\mathbf{X}^{(i)})$ for all $i \in [d]$ (see [29]). Thus, we can define an ideal $J_{d,HOSVD}$ 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 $\mathbf{r}_{TT} = (r_1, r_2, \dots, r_{d-1})$ such that $r_i = \text{rank}(\mathbf{X}^{(1, \dots, i)})$, $i \in [d-1]$ (see [49] for details). By taking into account only minors of order two of the matricizations $\tau \in \{\{1\}, \{1, 2\}, \dots, \{1, 2, \dots, d-1\}\}$, one may introduce a corresponding polynomial ideal $J_{d,TT}$.

Theorem 6 The polynomial ideals J_d , $J_{d,HOSVD}$, and $J_{d,TT}$ are equal, for all $d \geq 3$.

Proof Let $\tau \subset [d]$ represent a matricization. Similarly to the case of order-three tensors, for $(\alpha, \beta) \in \mathbb{N}^{2d}$, $x_{\alpha^\tau} x_{\beta^\tau}$ denotes the monomial where $\alpha_k^\tau = \alpha_k$, $\beta_k^\tau = \beta_k$ for all $k \in \tau$ and $\alpha_\ell^\tau = \beta_\ell^\tau = \beta_\ell$ for all $\ell \in \tau^c = [d] \setminus \tau$. Moreover, $x_{\alpha^{\tau,0}} x_{\beta^{\tau,0}}$ denotes the monomial where $\alpha_k^{\tau,0} = \alpha_k$, $\beta_k^{\tau,0} = \beta_k$ for all $k \in \tau$ and $\alpha_\ell^{\tau,0} = \beta_\ell^{\tau,0} = 0$ for all $\ell \in \tau^c = [d] \setminus \tau$. The corresponding order-two minors are defined as

$$f_{(\alpha,\beta)}^\tau(\mathbf{x}) = -x_\alpha x_\beta + x_{\alpha^\tau} x_{\beta^\tau}, \quad (\alpha, \beta) \in \mathcal{T}^\tau.$$

We define the set $\mathcal{T}^\tau$ as

$$\mathcal{T}^\tau = \left\{ (\alpha, \beta) : \alpha^{\tau,0} \neq \beta^{\tau,0}, \alpha^{\tau^c,0} \neq \beta^{\tau^c,0} \right\}.$$

Similarly as in the case of order-three tensors, notice that $f_{(\alpha,\beta)}^\tau(\mathbf{x}) = f_{(\beta,\alpha)}^\tau(\mathbf{x}) = -f_{(\alpha^\tau,\beta^\tau)}^\tau(\mathbf{x}) = -f_{(\beta^\tau,\alpha^\tau)}^\tau(\mathbf{x})$, for all $(\alpha, \beta) \in \mathcal{T}^\tau$. First, we show that $J_d = J_{d,\text{HOSVD}}$ by showing that $f_{(\alpha,\beta)}^\tau(\mathbf{x}) \in J_{d,\text{HOSVD}}$, for all $(\alpha, \beta) \in \mathcal{T}^\tau$ 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 $\mathcal{T}^\tau$, there exists at least one $\ell \in \tau^c$ such that $\alpha_\ell \neq \beta_\ell$. Let $\tau = \{t_1, t_2, \dots, t_k\}$ with $t_i < t_{i+1}$, for all $i \in [k-1]$ and $k \geq 2$. Next, fix $(\alpha, \beta) \in \mathcal{T}^\tau$ and define $\alpha^0 = \alpha$ and $\beta^0 = \beta$. Algorithm 3 results in polynomials $g_k \in J_{3,\text{TT}}$ such that $f_{(\alpha,\beta)}^\tau(\mathbf{x}) = \sum_{i=1}^k g_i(\mathbf{x})$. This follows from

$$\sum_{i=1}^k g_i = \sum_{i=1}^k (-x_{\alpha^{i-1}} x_{\beta^{i-1}} + x_{\alpha^i} x_{\beta^i}) = -x_{\alpha^0} x_{\beta^0} + x_{\alpha^k} x_{\beta^k} = f_{(\alpha,\beta)}^\tau(\mathbf{x}).$$

By the definition of polynomials g_k it is obvious that

$$g_i \in \left\{ f_{(\alpha,\beta)}^{[i]}(\mathbf{x}) : (\alpha, \beta) \in \mathcal{T}^{[i]} \right\}, \text{ for all } i \in [k].$$

Algorithm 3 Algorithm for proving that $J_d = J_{d,\text{TT}}$.

Input: An ideal $J_{d,\text{TT}} \in \mathbb{R}[\mathbf{x}]$, polynomial $f_{(\alpha,\beta)}^\tau(\mathbf{x})$ with $\alpha^0 = \alpha, \beta^0 = \beta, \tau = \{t_1, t_2, \dots, t_k\}$, where $k \geq 2$

for $i = 1, \dots, k$ **do**

Define α^i **and** β^i **as**

$$\alpha_j^i := \begin{cases} \beta_j^{i-1} & \text{if } j = t_i, \\ \alpha_j^{i-1} & \text{otherwise} \end{cases} \quad \text{and} \quad \beta_j^i := \begin{cases} \alpha_j^{i-1} & \text{if } j = t_i, \\ \beta_j^{i-1} & \text{otherwise.} \end{cases}$$

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

end for

Output: Polynomials $g_1, g_2, \dots, g_k$.

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

$$g(\mathbf{x}) = -x_\alpha x_\beta + x_{\alpha^{\{1,2,\dots,k\}}} x_{\beta^{\{1,2,\dots,k\}}}$$

$$h(\mathbf{x}) = -x_{\alpha^{\{1,2,\dots,k\}}} x_{\beta^{\{1,2,\dots,k\}}} + x_{\alpha^{\{1,2,\dots,k\}\{1,2,\dots,k-1\}}} x_{\beta^{\{1,2,\dots,k\}\{1,2,\dots,k-1\}}}$$

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

$$\begin{aligned} g_1(\mathbf{x}) &= -x_{\alpha}x_{\beta} + x_{\alpha\{1,2,\dots,k-1\}}x_{\beta\{1,2,\dots,k-1\}} \in \mathcal{R}^{\{1,2,\dots,k-1\}} \\ h_1(\mathbf{x}) &= -x_{\alpha\{1,2,\dots,k-1\}}x_{\beta\{1,2,\dots,k-1\}} + x_{\alpha\{1,2,\dots,k-1\}\{1,2,\dots,k\}}x_{\beta\{1,2,\dots,k-1\}\{1,2,\dots,k\}} \\ &= -x_{\alpha\{1,2,\dots,k\}}x_{\beta\{1,2,\dots,k\}} + x_{\alpha^{(k)}}x_{\beta^{(k)}} \in \mathcal{R}^{\{1,2,\dots,k\}}. \end{aligned} \quad \square$$

Remark 6 Fix a decomposition tree T_I which generates a particular HT-decomposition and consider the ideal J_{d,HT,T_I} generated by all second-order minors corresponding to the matricizations induced by the tree T_I . In a similar way as above, one can obtain that J_{d,HT,T_I} equals to J_d .

5 Convergence of the unit θ_k -norm balls

In this section we show the following result on the convergence of the unit θ_k -balls .

Theorem 7 *The theta body sequence of J_d converges asymptotically to the $\text{conv}(v_{\mathbb{R}}(J))$, i.e.,*

$$\bigcap_{k=1}^{\infty} TH_k(J_d) = \text{conv}(v_{\mathbb{R}}(J_d)).$$

To prove Theorem 7 we use the following result presented in [2] which is a consequence of Schmüdgen's Positivstellensatz.

Theorem 8 *Let J be an ideal such that $v_{\mathbb{R}}(J)$ is compact. Then the theta body sequence of J converges to the convex hull of the variety $v_{\mathbb{R}}(J)$, in the sense that*

$$\bigcap_{k=1}^{\infty} TH_k(J) = \text{conv}(v_{\mathbb{R}}(J)).$$

Proof of Theorem 7 The set $v_{\mathbb{R}}(J_d)$ is the set of rank-one tensors with unit Frobenius norm which can be written as $v_{\mathbb{R}}(J_d) = \mathcal{A}_1 \cap \mathcal{A}_2$ where

$$\begin{aligned} \mathcal{A}_1 &= \{\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times \dots \times n_d} : \text{rank}(\mathbf{X}) = 1\}, \\ \text{and } \mathcal{A}_2 &= \{\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times \dots \times n_d} : \|\mathbf{X}\|_F = 1\}. \end{aligned}$$

It is well-known that $\mathcal{A}_1$ is closed [11, discussion before Definition 2.2] and since $\mathcal{A}_2$ is clearly compact, $v_{\mathbb{R}}(J_d)$ is compact. Therefore, the result follows from Theorem 8. $\square$

6 Computational complexity

The computational complexity of the semidefinite programs for computing the θ_1 -norm of a tensor or for minimizing the θ_1 -norm subject to a linear constraint depends

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

Indeed, the moment matrix $\mathbf{M}$ is of dimension $(1 + n^d) \times (1 + n^d)$ (see also (16) for matrices in $\mathbb{R}^{2 \times 2}$) and if $a = n^d$ denotes the total number of entries of a tensor $\mathbf{X} \in \mathbb{R}^{n \times \dots \times n}$, then the number of the variables is at most $\frac{a(a+1)}{2} \sim \mathcal{O}(a^2)$ 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 θ_1 -norm minimization via numerical experiments, where we concentrate on third-order tensors. Due to large computation times with standard semidefinite solvers, we focus only on small tensors and leave the optimization of the algorithm for future work. Given measurements $\mathbf{b} = \mathcal{A}(\mathbf{X})$ of a low rank tensor $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, where $\mathcal{A} : \mathbb{R}^{n_1 \times n_2 \times n_3} \rightarrow \mathbb{R}^m$ is a linear measurement map, we aim at reconstructing $\mathbf{X}$ as the solution of the minimization program

$$\min \|\mathbf{Z}\|_{\theta_1} \quad \text{subject to } \mathcal{A}(\mathbf{Z}) = \mathbf{b}. \quad (23)$$

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

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

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

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

where $\ell = \sum_{r=2}^{p-1} |\mathbf{M}^r| + q$, $\mathbf{M}^p = \{\mathbf{M}_{\tilde{\ell}}^p\}$, and the matrices $\mathbf{M}_0$, $\mathbf{M}_{ijk}$ and $\mathbf{M}_{\tilde{\ell}}^p$ are provided in Table 3. For $p \in \{2, 3, \dots, 9\}$, the function h_p denotes an arbitrary but fixed bijection $\{1, 2, \dots, |\mathbf{M}^p|\} \mapsto \{(i, \hat{i}, j, \hat{j}, k, \hat{k})\}$, where $\tilde{\ell} = (i, \hat{i}, j, \hat{j}, k, \hat{k})$ is in the range of the last column of Table 3. As discussed in Section 2 for the general case, the θ_1 -norm minimization problem (23) is then equivalent to the semidefinite program

$$\min_{t, \mathbf{y}, \mathbf{Z}} t \quad \text{subject to} \quad \mathbf{M}(t, \mathbf{y}, \mathbf{Z}) \succeq 0 \quad \text{and} \quad \mathcal{A}(\mathbf{Z}) = \mathbf{b}. \quad (24)$$

For our experiments, the linear mapping is defined as $(\mathcal{A}(\mathbf{X}))_k = \langle \mathbf{X}, \mathcal{A}_k \rangle$, $k \in [m]$, with independent Gaussian random tensors $\mathcal{A}_k \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, i.e., all entries of $\mathcal{A}_k$ are independent $\mathcal{N}\left(0, \frac{1}{m}\right)$ random variables. We choose tensors $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ of rank one as $\mathbf{X} = \mathbf{u} \otimes \mathbf{v} \otimes \mathbf{w}$, where each entry of the vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ is

Table 3 The matrices involved in the definition of the moment matrix $\mathbf{M}(t, \mathbf{y}, \mathbf{X})$. Due to the symmetry only the upper triangle part of the matrices is specified

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

The other non-specified entries of the matrices $\mathbf{M} \in \mathbb{R}^{(n_1 n_2 n_3 + 1) \times (n_1 n_2 n_3 + 1)}$ from the first column are equal to zero. The matrix $\mathbf{M}$ corresponds to the element $g + J_3$ of the θ -basis specified in the second column. The index $\tilde{I} = (i, \hat{i}, j, \hat{j}, k, \hat{k})$ is in the range of the last column. The function $f: \mathbb{N}^3 \rightarrow \mathbb{N}$ is defined as $f(i, j, k) = (i - 1)n_2 n_3 + (j - 1)n_3 + k + 1$

taken independently from the normal distribution $\mathcal{N}(0, 1)$. Tensors $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ of rank two are generated as the sum of two random rank-one tensors. With $\mathcal{A}$ and $\mathbf{X}$ given, we compute $\mathbf{b} = \mathcal{A}(\mathbf{X})$, run the semidefinite program (24) and compare its minimizer with the original low rank tensor $\mathbf{X}$. For a given set of parameters, i.e., dimensions n_1, n_2, n_3 , number of measurements m and rank r , we repeat this experiment 200 times and record the empirical success rate of recovering the original tensor, where we say that recovery is successful if the elementwise reconstruction error is at most 10^{-6} . We use MATLAB (R2008b) for these numerical experiments, including SeDuMi.1.3 for solving the semidefinite programs.

Table 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

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	$n_1 n_2 n_3$	cpu (s)
$2 \times 2 \times 3$	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

the last column. Alternatively, the SDPNAL+ MATLAB toolbox (version 0.5 beta) for semidefinite programming [62, 64] allows to perform low rank tensor recovery via θ_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 5 min). Similarly, rank-one $11 \times 11 \times 11$ tensors are recovered from $m = 125$ measurements with one simulation lasting about 50 min. 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.

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 θ_1 -minimization seems to be a promising approach. Of course, it remains to investigate the recovery performance theoretically.

Figures 1 and 2 present the numerical results for low rank tensor recovery via θ_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 elementwise reconstruction error is smaller than 10^{-3} . Figures 1a, 2a, and 3a and 1b, 2b, and 3b present experiments for rank-one and rank-two tensors, respectively. The vertical axis in all three figures represents the empirical success rate. In Fig. 1 the horizontal axis represents the relative number of measurements, to be more precise, for a tensor of size $n \times n \times n$, the number $\bar{n}$ on the horizontal axis represents $m = \bar{n} \frac{n^3}{100}$ measurements. In Fig. 2 for a rank- r tensor of size $n \times n \times n$ and the number of measurements m , the horizontal axis represents the number $m/(3nr)$. Notice that $3nr$ represents the degrees of freedom in the corresponding CP-decomposition. In particular, if the number of measurements necessary for tensor recovery is $m \geq 3Cnr$, for an universal constant C , Fig. 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 $m \geq 3Cnr$ 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 Fig. 3 represents

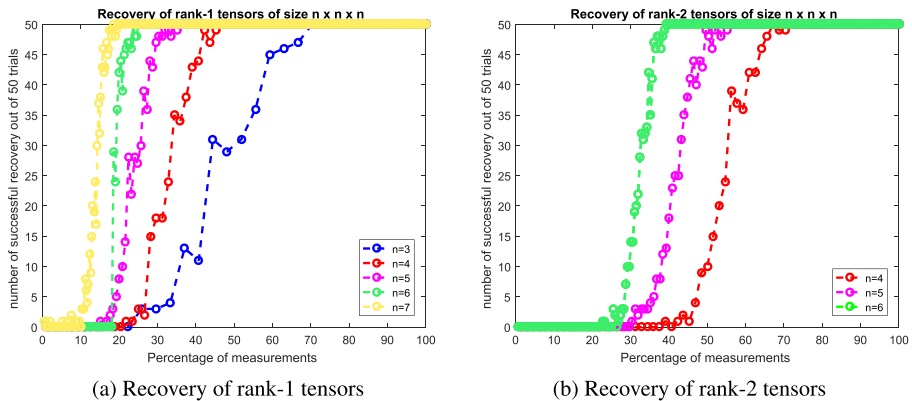


Fig. 1 Recovery of rank-1 and rank-2 tensors via θ_1 -norm minimization

the number $m/(3nr \cdot \log(n))$. The figure suggests that with the number of measurements $m \geq 6rn \cdot \log(n)$ we would always be able to recover a low rank tensor and therefore it may be possible that a logarithmic factor is necessary.

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 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.

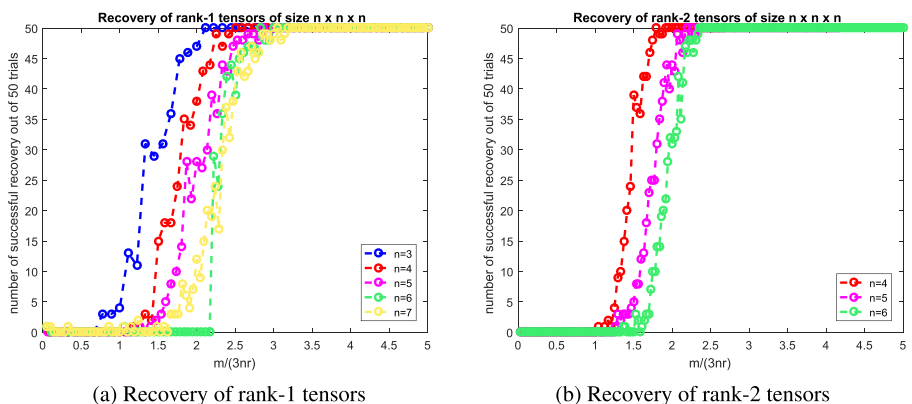


Fig. 2 Recovery of rank-1 and rank-2 tensors via θ_1 -norm minimization

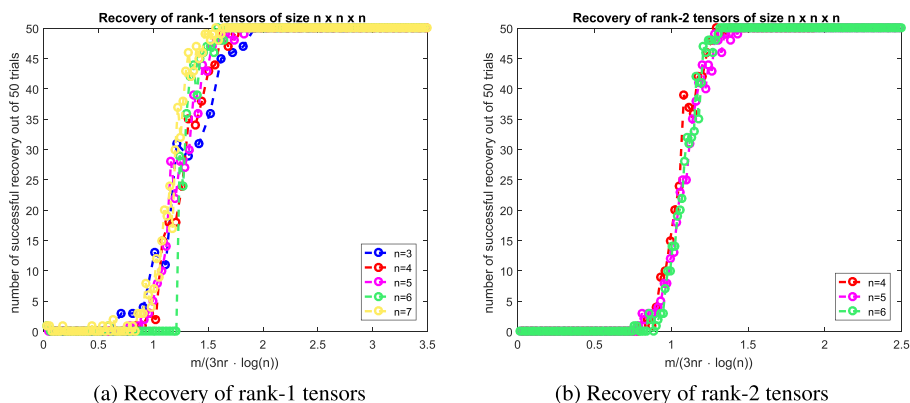


Fig. 3 Recovery of rank-1 and rank-2 tensors via θ_1 -norm minimization

Acknowledgments 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.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix Monomial orderings and Gröbner bases

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

Definition 7 For $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)$, $\beta = (\beta_1, \beta_2, \dots, \beta_n) \in \mathbb{Z}_{\geq 0}^n$, we write $\mathbf{x}^\alpha >_{\text{grevlex}} \mathbf{x}^\beta$ (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 [15, 16].

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 8 (Gröbner basis) For a fixed monomial order, a basis $\mathcal{G} = \{g_1, g_2, \dots, g_s\}$ of a polynomial ideal $J \subset \mathbb{R}[\mathbf{x}]$ is a *Gröbner basis* (or standard basis) if for all $f \in \mathbb{R}[\mathbf{x}]$ there exist a **unique** $r \in \mathbb{R}[\mathbf{x}]$ and $g \in J$ such that $f = g + r$ and no monomial of r is divisible by any of the leading monomials in $\mathcal{G}$, i.e., by any of the monomials $LM(g_1), LM(g_2), \dots, LM(g_s)$.

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

Definition 9 The *reduced Gröbner basis* for a polynomial ideal $J \in \mathbb{R}[\mathbf{x}]$ is a Gröbner basis $\mathcal{G} = \{g_1, g_2, \dots, g_s\}$ for J such that

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

In other words, a Gröbner basis $\mathcal{G}$ is the reduced Gröbner basis if for all $i \in [s]$ the polynomial $g_i \in \mathcal{G}$ is monic (i.e., $LC(g_i) = 1$) and the leading monomial $LM(g_i)$ 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 8 equals zero. Gröbner bases are also one of the main computational tools in solving systems of polynomial equations [16].

With $\overline{f}^F$ we denote the remainder on division of f by the ordered k -tuple $F = (f_1, f_2, \dots, f_k)$. If F is a Gröbner basis for an ideal $\langle f_1, f_2, \dots, f_k \rangle$, then we can regard F as a set without any particular order by Definition 8, or in other words, the result of the division algorithm does not depend on the order of the polynomials. Therefore, $\overline{f}^{\mathcal{G}} = r$ in Definition 8.

The following result follows directly from Definition 8 and the polynomial division algorithm [16].

Corollary 2 Fix a monomial ordering and let $\mathcal{G} = \{g_1, g_2, \dots, g_s\} \subset \mathbb{R}[\mathbf{x}]$ be a Gröbner basis of a polynomial ideal J . A polynomial $f \in \mathbb{R}[\mathbf{x}]$ is in the ideal J if it can be written in the form $f = a_1g_1 + a_2g_2 + \dots + a_sg_s$, where $a_i \in \mathbb{R}[\mathbf{x}]$, for all $i \in [s]$, s.t. whenever $a_i g_i \neq 0$ we have

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

Definition 10 Fix a monomial order and let $\mathcal{G} = \{g_1, g_2, \dots, g_s\} \subset \mathbb{R}[\mathbf{x}]$. Given $f \in \mathbb{R}[\mathbf{x}]$, we say that f reduces to zero modulo $\mathcal{G}$ and write

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

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

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

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

The Gröbner basis of a polynomial ideal always exists and can be computed in a finite number of steps via Buchberger's algorithm [5, 15, 16].

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

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

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

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

Theorem 9 (Buchberger's criterion) *A basis $\mathcal{G} = \{g_1, g_2, \dots, g_s\}$ for a polynomial ideal $J \subset \mathbb{R}[\mathbf{x}]$ is a Gröbner basis if and only if $S(g_i, g_j) \rightarrow_{\mathcal{G}} 0$ for all $i \neq j$.*

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

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

$$LCM(LM(f), LM(g)) = LM(f)LM(g),$$

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

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

References

1. Bhatia, R.: Matrix analysis. Graduate texts in mathematics. vol. 169, Springer (1996)

2. Blekherman, G., Parrilo, P.A., Thomas, R.R.: Semidefinite optimization and convex algebraic geometry SIAM (2013)
3. Boyd, S., Vandenberghe, L.: Convex optimization. Cambridge univ press (2004)
4. Brylinski, J.-L.: Algebraic Measures of Entanglement. In: Chen, G., Brylinski, R. K., Mathematics of Quantum Computation. CRC, Boca Raton, FL (2002)
5. Buchberger, B.: 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. Candès, E.J., Plan, Y.: 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. Candès, E.J., Recht, B.: Exact matrix completion via convex optimization. Found. Comput. Math. **9**(6), 717–772 (2009)
8. Candès, E.J., Strohmer, T., Vershyninski, V.: PhaseLift: exact and stable signal recovery from magnitude measurements via convex programming. Comm. Pure Appl. Math. **66**(8), 1241–1274 (2013). <https://doi.org/10.1002/cpa.21432>
9. Candès, E.J., Tao, T.: The power of matrix completion: near-optimal convex relaxation. IEEE Trans Information Theory **56**(5), 2053–2080 (2010)
10. Candès, E.J., Tao, T., Romberg, J.K.: Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Trans. Inform. Theory **52**(2), 489–509 (2006)
11. Cartwright, D., Erman, D., Oeding, L.: Secant varieties of $\mathbb{P}^2 \times \mathbb{P}^n$ embedded by $\mathcal{O}(1, 2)$. J. London Math. Soc. **85**(1), 121–141 (2012)
12. Chandrasekaran, V., Recht, B., Parrilo, P.A., Willsky, A.: The convex geometry of linear inverse problems. Found. Comput. Math. **12**(6), 805–849 (2012)
13. Chen, Y., Bhojanapalli, S., Sanghavi, S., Ward, R.: Completing any low-rank matrix, provably. J. Mach. Learn. Res. **16**, 2999–3034 (2015)
14. Combettes, P., Pesquet, J.C., Proximal Splitting Methods in Signal Processing. In: H. Bauschke, R. Burachik, P. Combettes, V. Elser, D. Luke, H. Wolkowicz (Eds.) Fixed-Point Algorithms for Inverse Problems in Science and Engineering, pp. 185–212. Springer (2011)
15. Cox, D., Little, J., O'Shea, D. Using Algebraic Geometry. Graduate Texts in Mathematics, Second edn, vol. 185. Springer, New York (2005)
16. Cox, D., Little, J., O'Shea, D. Ideals, Varieties, and Algorithms, Third edn. Undergraduate Texts in Mathematics. Springer, New York (2007)
17. Da Silva, C., Herrmann, F.J.: Hierarchical Tucker Tensor Optimization-Applications to Tensor Completion. In: SAMPTA 2013, pp. 384–387 (2013)
18. De Silva, V., Lim, L.-H.: Tensor rank and ill-posedness of the best low-rank approximation problem. SIAM J. Matrix Anal. Appl. **30**(3), 1084–1127 (2008)
19. Defant, A., Floret, K.: Tensor norms and operator ideals. North-holland mathematics studies elsevier science (1992)
20. Donoho, D.L.: Compressed sensing. IEEE Trans. Inform. Theory **52**(4), 1289–1306 (2006)
21. Duarte, M.F., Baraniuk, R.G.: Kronecker compressive sensing. IEEE Trans Image Proc (2011)
22. Fazel, M.: Matrix rank minimization with applications. Ph.D thesis (2002)
23. Foucart, S., Rauhut, H.: A Mathematical Introduction to Compressive Sensing, Applied and Numerical Harmonic Analysis birkhäuser (2013)
24. Friedland, S., Lim, L.-H.: Nuclear norm of higher-order tensors. Math. Comp. **87**(311), 1255–1281 (2018)
25. Gandy, S., Recht, B., Yamada, I.: Tensor completion and low- n -rank tensor recovery via convex optimization. Inverse Problems **27**(2), 025010 (2011)
26. Gouveia, J., Laurent, M., Parrilo, P.A., Thomas, R.R.: A new semidefinite programming hierarchy for cycles in binary matroids and cuts in graphs, Math. Prog., 1–23 (2009)
27. Gouveia, J., Parrilo, P.A., Thomas, R.R.: Theta bodies for polynomial ideals. SIAM J. Optim. **20**(4), 2097–2118 (2010)
28. Grande, F., Sanyal, R.: Theta rank, levelness, and matroid minors. J Combin. Theory Ser. B **127**, 1–31 (2017)
29. Grasedyck, L.: Hierarchical singular value decomposition of tensors. SIAM J. Matrix Anal. Appl **31**, 2029 (2010)
30. Grasedyck, L., Hackbusch, W.: An introduction to hierarchical (H-) rank and TT-rank of tensors with examples. Comput. Methods Appl. Math. **11**(3), 291–304 (2011)

31. Gross, D.: Recovering low-rank matrices from few coefficients in any basis. *IEEE Trans Inform. Theory* **57**(3), 1548–1566 (2011)
32. Gross, D., Liu, Y.-K., Flammia, S.T., Becker, S., Eisert, J.: Quantum state tomography via compressed sensing. *Phys. Rev. Lett.* **150****401**, 105 (2010)
33. Hackbusch, W.: *Tensor spaces and numerical tensor calculus* Springer (2012)
34. Hibi, T.: 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)
35. Hillar, C.J., Lim, L.-H.: Most tensor problems are NP-hard. *J. ACM* **60**(6), 45, 1–45, 39 (2013)
36. Hitchcock, F.L.: The expression of a tensor or a polyadic as a sum of products. *J. Math. Phys.* **6**(1–4), 164–189 (1927). <https://doi.org/10.1002/sapm192761164>
37. Hitchcock, F.L.: Multiple invariants and generalized rank of a p-way matrix or tensor. *J. Math. Phys.* **7**(1), 39–79 (1927)
38. Håstad, J.: Tensor rank is nP-complete. *J. Algorithms* **11**(4), 644–654 (1990)
39. Huang, B., Mu, C., Goldfarb, D., Wright, J.: Provable models for robust low-rank tensor recovery. *Pac. J. Optim* **11**(2), 339–364 (2015)
40. Karlsson, L., Kressner, D., Uschmajew, A.: Parallel algorithms for tensor completion in the CP format. *Parallel Comput.* **57**, 222–234 (2016). <https://doi.org/10.1016/j.parco.2015.10.002>
41. Kreimer, N., Sacchi, M.: A tensor higher-order singular value decomposition for prestack seismic data noise reduction and interpolation. *Geophys. J. Internat.* **77** v113–V122 (2012)
42. Kressner, D., Steinlechner, M., Vandereycken, B.: Low-rank tensor completion by Riemannian optimization. *BIT Numer Math.* **54**(2), 447–468 (2014)
43. Kueng, R., Rauhut, H., Terstiege, U.: Low rank matrix recovery from rank one measurements. *Appl. Comput. Harmon. Anal.* **42**(1), 88–116 (2017)
44. Landsberg, J.M.: *Tensors: Geometry and Applications*. Graduate studies in mathematics American Mathematical Society (2011)
45. Lasserre, J.: *Moments, Positive Polynomials and Their Applications* Imperial College Press Optimization Series, 1, Imperial College Press, London (2010)
46. Liu, J., Musialski, P., Wonka, P., Ye, J.: Tensor Completion for Estimating Missing Values in Visual Data. In: *IC 1V* (2009)
47. Liu, Y., Shang, F., Fan, W., Cheng, J., Cheng, H.: Generalized Higher-Order Orthogonal Iteration for Tensor Decomposition and Completion. In: *Advances in Neural Information Processing Systems*, pp. 1763–1771 (2014)
48. Lovász, L.: On the Shannon capacity of a graph. *IEEE Trans Inform. Theory* **25**(1), 1–7 (1979)
49. Oseledets, I.V.: Tensor-train decomposition. *SIAM J. Sci. Comput.* **33**(5), 2295–2317 (2011)
50. Oymak, S., Jalali, A., Fazel, M., Eldar, Y.C., Hassibi, B.: Simultaneously structured models with application to sparse and low-rank matrices. *IEEE Trans. Inform. Theory* **61**(5), 2886–2908 (2015). <https://doi.org/10.1109/TIT.2015.2401574>
51. Parikh, N., Boyd, S.: Proximal algorithms. *Found. Trends Optim.* **1**(3), 123–231 (2013)
52. Rauhut, H., Schneider, R., Stojanac, Ž.: Tensor Tensor Recovery via Iterative Hard Thresholding. In: *Proc. SampTA 2013* (2013)
53. Rauhut, H., Schneider, R., Stojanac, Ž.: Tensor Completion in Hierarchical Tensor Representations. In: H. Boche, R. Calderbank, G. Kutyniok, J. Vybiral (Eds.) *Compressed Sensing and Its Applications*. Springer (2015)
54. Rauhut, H., Schneider, R., Stojanac, Ž.: Low rank tensor recovery via iterative hard thresholding. *Linear Algebra Appl.* **523**, 220–262 (2017)
55. Recht, B., Fazel, M., Parrilo, P.: Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Rev.* **52**(3), 471–501 (2010)
56. Romera-Paredes, B., Aung, H., Bianchi-Berthouze, N., Pontil, M.: Multilinear multitask learning. *J. Mach. Learn. Res.* **28**(3), 1444–1452 (2013)
57. Ryan, R.A.: *Introduction to tensor products of banach spaces*. Celtic studies springer (2002)
58. Stojanac, Ž.: *Low-Rank Tensor Recovery*, Ph.D. thesis, Universität Bonn (2016)
59. Toh, K., Yun, S.: An accelerated proximal gradient algorithm for nuclear norm regularized least squares problems. *Pac. J. Optim.* **6**, 615–640 (2010)
60. Tomioka, R., Hayashi, K. arXiv:1010.0789 (2010)
61. Wong, Y.-C.: *Schwartz spaces, nuclear spaces, and tensor products*. Lecture notes in mathematics Springer-Verlag (1979)

62. Yang, L.Q., Sun, D.F., Toh, K.C.: SDPNAL+: A majorized semismooth newton-CG augmented Lagrangian method for semidefinite programming with nonnegative constraints. *Math. Program. Comput.* **7**(3), 331–366 (2015)
63. Yuan, M., Zhang, C.-H.: on tensor completion via nuclear norm minimization. *Found. Comput. Math.* **16**(4), 1031–1068 (2016). <https://doi.org/10.1007/s10208-015-9269-5>
64. Zhao, X.Y., Sun, D.F., Toh, K.C.: A newton-CG Augmented Lagrangian Method for Semidefinite Programming. *SIAM. J. Optimization* **20**(4), 1737–1765 (2010)

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Holger Rauhut¹  · Željka Stojanac¹

Željka Stojanac
zeljka.stojanac@gmail.com

¹ Chair for Mathematics of Information Processing, RWTH Aachen University, Pontdriesch 10, 52056 Aachen, Germany