Overcomplete Tensor Decomposition via Convex Optimization

Qiuwei Li*, Ashley Prater†, Lixin Shen‡, and Gongguo Tang*
*Department of Electrical Engineering and Computer Science,
Colorado School of Mines, Golden, CO 80401 Email: qiuli@mymail.mines.edu, gtang@mines.edu
† Air Force Research Laboratory, Rome, NY 13441 Email: ashley.prater.3@us.af.mil
‡Department of Mathematics, Syracuse University, Syracuse, NY 13244 Email: lshen03@syr.edu

***Abstract*—This work develops theories and computational methods for *overcomplete, non-orthogonal* tensor decomposition using convex optimization. Under an incoherence condition of the rank-one factors, we show that one can retrieve tensor decomposition by solving a convex, infinite-dimensional analog of ℓ_1 minimization on the space of measures. The optimal value of this optimization defines the tensor nuclear norm. Two computational schemes are proposed to solve the infinite-dimensional optimization: semidefinite programs based on sum-of-squares relaxations and nonlinear programs that are an exact reformulation of the tensor nuclear norm. The latter exhibits superior performance compared with the state-of-the-art tensor decomposition methods.**

I. Introduction

Tensors provide natural representations for massive multi-mode datasets encountered in image and video processing [1], collaborative filtering [2], array signal processing [3], and psychometrics [4]. Tensor methods also form the backbone of many machine learning, signal processing, and statistical algorithms, including independent component analysis (ICA), latent graphical model learning [5], dictionary learning [6], and Gaussian mixture estimation. The utility of tensors in such diverse applications is mainly due to the ability to identify *overcomplete*, *non-orthogonal* factors from tensor data as already suggested by the celebrated Kruskal's theorem. This is in sharp contrast to the inherent ambiguous nature of matrix decompositions.

Among several important tensor inverse problems, including decomposition, completion, approximation, and denoising, tensor decomposition is the most fundamental one as it lays out theoretical foundations that solutions of other problems can build upon. In this work, we frame tensor decomposition as a sparse recovery problem where the underlying dictionary is parameterized by points on the unit spheres. More precisely, we consider the following decomposition model:

$$T = \sum\nolimits_{p=1}^{r} \lambda_p \mathbf{u}_p \otimes \mathbf{v}_p \otimes \mathbf{w}_p, \tag{1}$$

where $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{S}^{n-1}$, the unit sphere of $\mathbb{R}^n$, and $\otimes$ is the tensor/outer product defined via $[\mathbf{u} \otimes \mathbf{v}]_{i,j} = u_i v_j$. The decomposition with the smallest r is called a Canonical Polyadic (CP) decomposition and the corresponding r is the rank of the tensor T. When the set of rank-one tensors $\mathcal{A} = \{\mathbf{u} \otimes \mathbf{v} \otimes \mathbf{w} : \mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{S}^{n-1}\}$ is viewed as a dictionary with infinite number of atoms, the CP decomposition finds the sparsest representation of T with respect to this dictionary. Overcomplete decomposition refers to the regime that the rank r is greater than the dimension n. Inspired by the success of ℓ_1 minimization in finding sparse representations for finite dictionaries [7], we employ an analog infinite-dimensional generalization for tensor decomposition. The major theoretical problem we investigate is: under what conditions on the rank-one factors, the ℓ_1 minimization returns the CP decomposition?

Despite the advantages provided by tensor methods in many applications, their widespread adoption has been slow due to inherent computational intractability. Although the CP decomposition (1) is a multi-mode generalization of the singular value decomposition for matrices, determining the CP decomposition for a given tensor is a non-trivial problem that is still under active investigation (cf. [8], [9]). Indeed, even determining the rank of a third-order tensor is an NP-hard problem [10]. A common strategy used to compute a tensor decomposition is to apply an alternating minimization scheme. Although efficient, this approach has the drawback of not providing global convergence guarantees [8]. Recently, an approach combining alternating minimization with power iteration has gained popularity due to its ability to guarantee tensor decomposition results under certain assumptions [11], [12]. Another closely related work [13] employs the same atomic norm idea but focuses on symmetric tensors. In addition, the result of [13] does not apply to overcomplete decompositions.

The paper is organized as follows. Section II sets up the signal model and formulates the problem to be solved. In Section III, we present the main theorem and outline its proof. Section IV presents two computational methods to solve the tensor decomposition. Section V is devoted to numerical experiments and Section VI concludes the paper.

II. Signal Model and Problem Setup

In this section, we view tensor decomposition as the estimation of a measure from its moments, and formulate the optimization we will use to extract the decomposition.

Define a Borel (non-negative) measure $\mu = \sum_{p=1}^{r} \lambda_p \delta(\mathbf{u} - \mathbf{u}_p, \mathbf{v} - \mathbf{v}_p, \mathbf{w} - \mathbf{w}_p)$ on the set $\mathbb{K} = \mathbb{S}^{n-1} \times \mathbb{S}^{n-1} \times \mathbb{S}^{n-1}$. Here $\delta(\mathbf{u} - \mathbf{u}_p, \mathbf{v} - \mathbf{v}_p, \mathbf{w} - \mathbf{w}_p)$ represents the Dirac measure whose total, unit mass concentrates on a single point $(\mathbf{u}_p, \mathbf{v}_p, \mathbf{w}_p) \in \mathbb{K}$. Apparently, the measure μ is a discrete measure on $\mathbb{K}$ that is supported on a finite number of points. This perspective allows us to write any decomposition of the form (1) as $T = \int_{\mathbb{K}} \mathbf{u} \otimes \mathbf{v} \otimes \mathbf{w} d\mu$.

L. Shen was supported by the NSF under Awards DMS-1522332 and DMS-1115523. Q. Li and G. Tang were supported by the NSF Grant CCF-1464205. G. Tang was also supported by the 2015 Air Force Summer Faculty Fellowship Program and the Air Force Summer Research Extension Grant.

Approved for public release by WPAFB public affairs on 15 Jul 2015. Case number 88ABW-2015-3608.

978-1-4799-1963-5/15/$31.00 ©2015 IEEE

Suppose the following true decomposition is unknown:

$$T = \sum_{p=1}^{r} \lambda_p^\star \mathbf{u}_p^\star \otimes \mathbf{v}_p^\star \otimes \mathbf{w}_p^\star := \int_{\mathbb{K}} \mathbf{u} \otimes \mathbf{v} \otimes \mathbf{w} d\mu^\star. \quad (2)$$

To determine the decomposition from the tensor T, or equivalently, to recover the measure $\mu^\star = \sum_{p=1}^{r} \lambda_p^\star \delta(\mathbf{u} - \mathbf{u}_p^\star, \mathbf{v} - \mathbf{v}_p^\star, \mathbf{w} - \mathbf{w}_p^\star)$, we propose solving the optimization

$$\underset{\mu \in \mathcal{M}(\mathbb{K})}{\text{minimize}}\, \mu(\mathbb{K}) \text{ subject to } T = \int_{\mathbb{K}} \mathbf{u} \otimes \mathbf{v} \otimes \mathbf{w} d\mu, \quad (3)$$

where $\mathcal{M}(\mathbb{K})$ is the set of Borel measures on $\mathbb{K}$. The optimal value of (3) defines $\|T\|_*$, the tensor nuclear norm, which is a special case of atomic norms [14] corresponding to the atomic set $\mathcal{A} = \{\mathbf{u} \otimes \mathbf{v} \otimes \mathbf{w} : \mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{S}^{n-1}\}$.

In this work, we establish conditions under which the decomposition given in (2) is the one that achieves the minimum, i.e., $\|T\|_* = \sum_{p=1}^{r} \lambda_p^\star$.

III. Main Theorem

In this section, we present the main theorem characterizing the exact recovery performance of the optimization (3). We first heuristically argue conditions the tensor should satisfy in order to solve (3). Suppose two rank-one factors $(\mathbf{u}_1, \mathbf{v}_1, \mathbf{w}_1) \approx (\mathbf{u}_2, \mathbf{v}_2, \mathbf{w}_2)$, then $\mathbf{u}_1 \otimes \mathbf{v}_1 \otimes \mathbf{w}_1$ and $\mathbf{u}_2 \otimes \mathbf{v}_2 \otimes \mathbf{w}_2$ are also close, making the decomposition not economical when measured by the ℓ_1 norm. Therefore, such a decomposition is unlikely to be optimal to (3), suggesting that correct recovery depends on the incoherence [15] of the rank-one factors:

$$\Delta := \max_{p \neq q} \{\langle \mathbf{u}_p, \mathbf{u}_q \rangle, \langle \mathbf{v}_p, \mathbf{v}_q \rangle, \langle \mathbf{w}_p, \mathbf{w}_q \rangle\} \quad (4)$$

Indeed, [16] shows that a necessary condition for a decomposition to achieve the tensor nuclear norm is Δ smaller than a constant. We will need a stronger assumption:

Assumption I. The tensor factors are incoherent:

$$\max_{p \neq q} \{|\langle \mathbf{u}_p^\star, \mathbf{u}_q^\star \rangle|, |\langle \mathbf{v}_p^\star, \mathbf{v}_q^\star \rangle|, |\langle \mathbf{w}_p^\star, \mathbf{w}_q^\star \rangle|\} \leq \frac{\operatorname{polylog}(n)}{\sqrt{n}}. \quad (5)$$

The incoherence assumption at its current form is still not strong enough to ensure that $\mu^\star$ is the optimal solution of (3). We need the following two additional assumptions:

Assumption II. The spectral norms of $U = [u_1^\star \cdots u_r^\star]$, $V = [v_1^\star \cdots v_r^\star]$, $W = [w_1^\star \cdots w_r^\star]$ are well-controlled.

$$\max\{\|U\|, \|V\|, \|W\|\} \leq 1 + c\sqrt{\frac{r}{n}}$$

for some numerical constant c.

Assumption III. The Hadamard product of the Gram matrices of U and V satisfy an isometry condition:

$$\|(U'U) \odot (V'V) - I_r\| \leq \operatorname{polylog}(n) \frac{\sqrt{r}}{n}.$$

and similar bounds also hold for U, W, and V, W.

Now we are ready to present the main theorem of the paper:

Theorem 1: Under Assumptions I, II, and III, if in addition $r = O(n^{17/16} / \operatorname{polylog}(n))$, then for sufficiently large n, the decomposition in (2) uniquely solves (3).

Since $r = O(n^{17/16} / \operatorname{polylog}(n)) \gg n$, we obtain guaranteed *overcomplete* tensor decomposition via (3). We remark that if the incoherence bound in Assumption A is strengthened to $O(\frac{1}{n \operatorname{polylog}(n)})$, then the Assumptions II and III are consequences of Assumption I, implying Theorem 1 with possibly relaxed conditions on the rank. We prefer the current assumptions since they are satisfied with high probability if the factors $\{\mathbf{u}_p^\star\}, \{\mathbf{v}_p^\star\}$ and $\{\mathbf{w}_p^\star\}$ are generated according to uniform distributions on the spheres [11], leading to the following corollary:

Corollary 1: If the factors $\{\mathbf{u}_p^\star\}, \{\mathbf{v}_p^\star\}$ and $\{\mathbf{w}_p^\star\}$ are generated according to uniform distributions on the spheres, then for sufficiently large n with high probability the decomposition in (2) uniquely solves (3).

The proof of Theorem 1 relies on the construction of a dual certificate satisfying conditions specified in the following proposition:

Proposition 1: Suppose $S := \{\mathbf{u}_p^\star \otimes \mathbf{v}_p^\star \otimes \mathbf{w}_p^\star\}$ are linearly independent. If there exists a dual certificate $Q \in \mathbb{R}^{n \times n \times n}$ such that the corresponding dual polynomial

$$q(\mathbf{u}, \mathbf{v}, \mathbf{w}) := \langle Q, \mathbf{u} \otimes \mathbf{v} \otimes \mathbf{w} \rangle \quad (6)$$

satisfies the following Bounded Interpolation Conditions:

$$\begin{aligned} q(\mathbf{u}_p^\star, \mathbf{v}_p^\star, \mathbf{w}_p^\star) &= 1, p = 1, \ldots, r & \text{(Interpolation)} \quad (7) \\ q(\mathbf{u}, \mathbf{v}, \mathbf{w}) &< 1, \forall (\mathbf{u}, \mathbf{v}, \mathbf{w}) \in \mathbb{K} \setminus S & \text{(Boundedness)} \quad (8) \end{aligned}$$

then the decomposition (2) uniquely achieves $\|T\|_*$.

According to this proposition, to prove Theorem 1, it suffices to construct a dual polynomial of the form (6) that satisfies the Bounded Interpolation Conditions (7) and (8). The Boundedness condition (8) is hard to enforce directly. So we relax it to one that requires the dual polynomial $q(\mathbf{u}, \mathbf{v}, \mathbf{w})$ to achieve maximum on $\mathbb{K}$ at $(\mathbf{u}_p, \mathbf{v}_p, \mathbf{w}_p)$, a necessary condition for which is given as follows:

$$\sum_{j,k} Q_{ijk} \mathbf{v}_p^\star(j) \mathbf{w}_p^\star(k) = \mathbf{u}_p^\star(i), i \in [n], p \in [r]; \quad (9)$$

$$\sum_{i,k} Q_{ijk} \mathbf{u}_p^\star(i) \mathbf{w}_p^\star(k) = \mathbf{v}_p^\star(j), j \in [n], p \in [r]; \quad (10)$$

$$\sum_{i,j} Q_{ijk} \mathbf{u}_p^\star(i) \mathbf{v}_p^\star(j) = \mathbf{w}_p^\star(k), k \in [n], p \in [r]. \quad (11)$$

Here $[n] = \{1, 2, \ldots, n\}$ and $[r]$ is defined similarly. It's easy to verify that these conditions subsume the Interpolation condition (7). For notational simplicity, we represent the conditions in matrix vector form as $\mathcal{A}(Q) = \mathbf{s}$, where the definitions of $\mathcal{A}$ and $\mathbf{s}$ are apparent.

Among all possible Q satisfying (9), (10), and (11), we choose the one with minimal energy, i.e., the solution of

$$\text{minimize} \|Q\|_F^2 \text{ subject to } \mathcal{A}(Q) = \mathbf{s},$$

where $\|Q\|_F^2 := \sum_{i,j,k} Q_{ijk}^2$. Least-squares theory states that the optimal Q has the form

$$Q = \sum_{p=1}^{r} (\boldsymbol{\alpha}_p \otimes \mathbf{v}_p^\star \otimes \mathbf{w}_p^\star + \mathbf{u}_p^\star \otimes \boldsymbol{\beta}_p \otimes \mathbf{w}_p^\star + \mathbf{u}_p^\star \otimes \mathbf{v}_p^\star \otimes \boldsymbol{\gamma}_p),$$

where $(\boldsymbol{\alpha}_p, \boldsymbol{\beta}_p, \boldsymbol{\gamma}_p)$ are determined by the normal equation. With such a Q the dual polynomial takes the form of

$$q(\mathbf{u},\mathbf{v},\mathbf{w}) = \sum_{p=1}^{r}[\langle \boldsymbol{\alpha}_p, \mathbf{u}\rangle\langle \mathbf{v}_p^\star, \mathbf{v}\rangle\langle \mathbf{w}_p^\star, \mathbf{w}\rangle + \langle \mathbf{u}_p^\star, \mathbf{u}\rangle\langle \boldsymbol{\beta}_p, \mathbf{v}\rangle\langle \mathbf{w}_p^\star, \mathbf{w}\rangle + \langle \mathbf{u}_p^\star, \mathbf{u}\rangle\langle \mathbf{v}_p^\star, \mathbf{v}\rangle\langle \boldsymbol{\gamma}_p, \mathbf{w}\rangle].$$

We outline the rest of the proof showing that the constructed $q(\mathbf{u},\mathbf{v},\mathbf{w})$ indeed satisfies the Bounded Interpolation Conditions. The argument consists of three main steps:
1) Showing that $\boldsymbol{\alpha}_p \approx \frac{1}{3}\mathbf{u}_p, \boldsymbol{\beta}_p \approx \frac{1}{3}\mathbf{v}_p$ and $\boldsymbol{\gamma}_p \approx \frac{1}{3}\mathbf{w}_p$.
2) Showing $q(\mathbf{u},\mathbf{v},\mathbf{w}) < 1$ in regions near each $(\mathbf{u}_p, \mathbf{v}_p, \mathbf{w}_p)$.
3) Showing $q(\mathbf{u},\mathbf{v},\mathbf{w}) < 1$ in regions far away from any $(\mathbf{u}_p, \mathbf{v}_p, \mathbf{w}_p)$.
Due to space limitation, details of the arguments are omitted.

IV. Computational Methods

In this section, we present two computational schemes that either approximate or exactly solve the optimization (3).

A. Sum-of-Squares Relaxations

As a special moment problem, the optimization (3) can be approximated increasingly tight by the semidefinite programs in the Lasserre relaxation hierarchy [17]. The Lasserre hierarchy proposes that instead of optimizing with respect to the measure μ in (3), one can equivalently optimize the (infinite-dimensional) moment sequence corresponding to μ:

$$\mathbf{m} = [m_{\boldsymbol{\alpha}}] = \int_{\mathbb{K}} \boldsymbol{\xi}^{\boldsymbol{\alpha}} \mu(d\boldsymbol{\xi}).$$

Here the combined variable $\boldsymbol{\xi} = [\mathbf{u}' \ \mathbf{v}' \ \mathbf{w}']' \in \mathbb{R}^{3n}$, the multi-integer index $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_{3n})$, and the monomial $\boldsymbol{\xi}^{\boldsymbol{\alpha}} = \xi_1^{\alpha_1}\xi_2^{\alpha_2}\cdots\xi_{3n}^{\alpha_{3n}}$. To get a finite-dimensional relaxation, we truncate the infinite-dimensional moment sequence $\mathbf{m}$ to a finite-dimensional vector $\mathbf{m}_{2d}$ that includes moments up to order $2d$, *i.e.*, to retain moments $m_{\boldsymbol{\alpha}}$ with $|\boldsymbol{\alpha}| = \sum_{i=1}^{3n} \alpha_i \leq 2d$. Three sets of linear matrix inequalities should be satisfied for a vector $\mathbf{m}_{2d}$ to be the $2d$th order truncation of a moment sequence on $\mathbb{K}$. One linear matrix inequality states that the moment matrix

$$M_{2d}(\mathbf{m}_{2d}) := \int_{\mathbb{K}} \begin{bmatrix} 1 \\ \xi_1 \\ \vdots \\ \xi_{3n}^d \end{bmatrix} \begin{bmatrix} 1 \\ \xi_1 \\ \vdots \\ \xi_{3n}^d \end{bmatrix}' d\mu$$

is positive semidefinite. The notation suggests $M_{2d}(\mathbf{m}_{2d})$ is a (linear) function of the truncated moment vector $\mathbf{m}_{2d}$. In addition, since the tensor entries are third order moments of the measure, elements of $\mathbf{m}_{2d}$ corresponding to these moments are known when $d \geq 2$, giving rise to the second set of linear equations. The fact that μ is supported on $\mathbb{K}$ leads to the third set of linear constraints. Combined with the fact that the objective function $\mu(\mathbb{K}) = \int_{\mathbb{K}} 1 d\mu = \mathbf{m}_{2d}(1)$, the final relaxation is a semidefinite program. These relaxations are also called sum-of-squares (SOS) relaxations as in the dual formulation the truncation process is equivalent to replacing positive polynomials with sum-of-squares polynomials [18]. Apparently, increasing the relaxation order d yields tighter approximations to the original optimization (3).

B. Low-rank Factorization and Non-linear Reformulation

Despite the nice theoretical properties of the Lasserre relaxation hierarchy, one of its major drawbacks is its high computational complexity–the variable size of the $2d$th order relaxation is $O(n^{2d})$, making it prohibitive to run on large n without specialized semidefinite program solvers. In this subsection, we propose solving (3) via nonlinear programming as indicated by the following proposition:

Proposition 2: Suppose the decomposition that achieves the tensor nuclear norm $\|T\|_*$ involves r terms and $\tilde{r} \geq r$, then $\|T\|_*$ is equal to the optimal value of the following optimization:

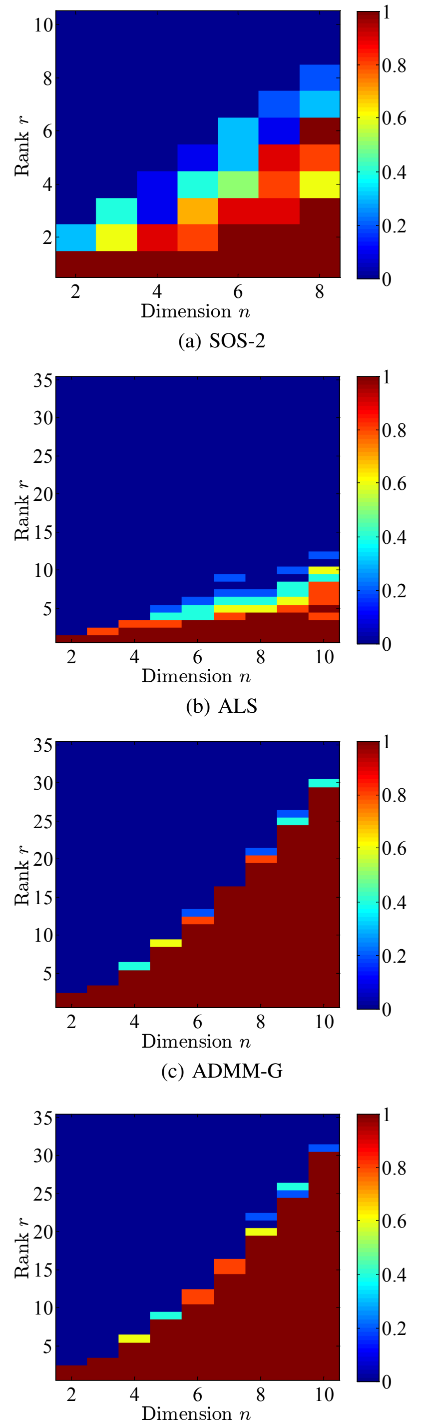
$$\begin{aligned} &\underset{U,V,W \in \mathbb{R}^{n\times\tilde{r}}}{\text{minimize}} \ \frac{1}{3}\left(\sum_{p=1}^{\tilde{r}}[\|U(:,p)\|_2^3 + \|V(:,p)\|_2^3 + \|W(:,p)\|_2^3]\right) \\ &\text{subject to } T = \sum_{p=1}^{\tilde{r}} U(:,p)\otimes V(:,p)\otimes W(:,p) \end{aligned} \tag{12}$$

where $U(:,p)$ denotes the pth column of the matrix U.

Therefore, when an upper bound on r is known, one can solve the nonlinear (and non-convex) program (12) to compute the tensor nuclear norm (and obtain the corresponding decomposition). Numerical simulations suggest that the nonlinear program (12), when solved using the ADMM approach, has superior performance. Particularly, although in theory only local optima can be obtained for the nonlinear programming formulation (12), in practice for tensors with randomly generated rank-one factors, the decomposition can always be recovered by the ADMM implementation of (12).

V. Numerical Experiments

We present numerical results to support the theory and to test the proposed computational methods. In particular, we examine the phase transition curves of the rate of success for four algorithms: i) the SOS relaxation of order $d = 2$ (SOS-2), ii) a state-of-the-art method based on power iteration initialization [11] (referred to as "Good Initialization") and alternating minimization refinement (ALS) [12], iii) ADMM implementation of (12) with "Good Initialization" (ADMM-G), and iv) ADMM with random initialization (ADMM-R). The phase transition curves are plotted in Figure 1. In preparing Figure 1, the r rank-one tensor components $\{(\mathbf{u}_p, \mathbf{v}_p, \mathbf{w}_p)\}_{p=1}^r$ were generated following i.i.d. Gaussian distribution, and then each $\mathbf{u}_p, \mathbf{v}_p, \mathbf{w}_p$ was normalized to have unit norm. We set the coefficients $\lambda_p = (1+\varepsilon_p^2)/2$, where ε_p is chosen from the standard normal distribution to ensure a minimal coefficient at least $1/2$. We varied the dimension n and the rank r. For each fixed (r, n) pair, 5 instances of the tensor were generated. We then run the four algorithms to each instance, and declared success if i) the recovered truncated moment vector is within 10^{-3} distance of the true moment vector for SOS method, and ii) the recovered tensor factors are within 10^{-3} distance to the true tensor factors. We use the moment vector criteria for SOS method because one cannot identify more than n tensor factors for the $d = 2$ relaxation. Also, by considering the high complexity of SOS method when n is large, we only set n from 2 to 8. The rate of success for each algorithm is the percentage of successful instances.



(a) SOS-2

(b) ALS

(c) ADMM-G

(d) ADMM-R

Fig. 1. Rate of success for tensor decomposition using SOS-2, ALS, ADMM-G, and ADMM-R.

From Figure 1, we observe that the SOS relaxation with $d = 2$ is unable to identify more than n factors. The ALS method with power iteration initialization also has difficulty in recovering more than n factors. On the other hand, the ADMM method works for r much larger than n. In addition, random initialization does not degrade the performance compared with "Good Initialization" using the power method of [11].

VI. Conclusions

By explicitly constructing a dual certificate, we derived conditions for a tensor decomposition to achieve the tensor nuclear norm. This implies that the infinite-dimensional measure optimization, which defines the tensor nuclear norm, is able to recover the decomposition under an incoherent condition and two other mild conditions. A computational method based on the Lasserre hierarchy and a nonlinear programming formulation were used to solve the measure optimization. Numerical experiments show that the nonlinear programming approach has superior performance. Future work will analyze the observed good performance of the nonlinear programming formulation.

References

[1] J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor Completion for Estimating Missing Values in Visual Data," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 1, pp. 208–220, 2013.

[2] A. Chen, "Context-Aware Collaborative Filtering System: Predicting the User's Preference in the Ubiquitous Computing Environment," in *Location- and Context-Awareness*, pp. 244–253, Berlin, Heidelberg: Springer Berlin Heidelberg, 2005.

[3] L.-H. Lim and P. Comon, "Multiarray signal processing: Tensor decomposition meets compressed sensing," *Comptes Rendus Mecanique*, vol. 338, no. 6, pp. 311–320, 2010.

[4] A. Smilde, R. Bro, and P. Geladi, *Multi-way Analysis: Applications in the Chemical Sciences*. John Wiley & Sons, 2005.

[5] A. Anandkumar, R. Ge, and M. Janzamin, "Analyzing tensor power method dynamics: Applications to learning overcomplete latent variable models," *arXiv preprint arXiv:1411.1488*, 2014.

[6] B. Barak, J. A. Kelner, and D. Steurer, "Dictionary learning and tensor decomposition via the sum-of-squares method," *arXiv preprint arXiv:1407.1543*, 2014.

[7] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via 1 minimization," *Proceedings of the National Academy of Sciences*, vol. 100, no. 5, pp. 2197–2202, 2003.

[8] P. Comon, "Tensor Decompositions, State of the Art and Applications," *Mathematics in Signal Processing V*, pp. 1–20, 2001.

[9] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Review*, vol. 51, no. 3, pp. 455–500, 2009.

[10] C. J. Hillar and L.-H. Lim, "Most Tensor Problems are NP-Hard," *Journal of the ACM (JACM)*, vol. 60, pp. 45–39, Nov. 2013.

[11] A. Anandkumar, R. Ge, and M. Janzamin, "Guaranteed non-orthogonal tensor decomposition via alternating rank-1 updates," *arXiv preprint arXiv:1402.5180*, 2014.

[12] P. Jain and S. Oh, "Provable tensor factorization with missing data," *Advances in Neural Information Processing Systems*, 2014.

[13] G. Tang and P. Shah, "Guaranteed Tensor Decomposition: A Moment Approach," in *International Conference on Machine Learning*, (Lille, France), 2015.

[14] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, "The Convex Geometry of Linear Inverse Problems," *Foundations of Computational Mathematics*, vol. 12, no. 6, pp. 805–849, 2012.

[15] L.-H. Lim and P. Comon, "Blind multilinear identification," *Information Theory, IEEE Transactions on*, vol. 60, no. 2, pp. 1260–1280, 2014.

[16] G. Tang, "Resolution limits for atomic decompositions via Markov-Bernstein inequalities ," in *Sampling Theory and Applications (SampTA), 2015 International Conference on*, pp. 548–552, IEEE, 2015.

[17] J.-B. Lasserre, *Moments, Positive Polynomials and Their Applications*. World Scientific, Oct. 2009.

[18] G. Blekherman, P. A. Parrilo, and R. R. Thomas, *Semidefinite Optimization and Convex Algebraic Geometry*. SIAM, Mar. 2013.