
Guaranteed Tensor Decomposition: A Moment Approach

Gongguo Tang

Colorado School of Mines, 1500 Illinois Street, Golden, CO, USA

GTANG@MINES.EDU

Parikshit Shah

Yahoo! Labs, 701 First Ave, Sunnyvale CA, USA

PSHAH@DISCOVERY.WISC.EDU

Abstract

We develop a theoretical and computational framework to perform guaranteed tensor decomposition, which also has the potential to accomplish other tensor tasks such as tensor completion and denoising. We formulate tensor decomposition as a problem of measure estimation from moments. By constructing a dual polynomial, we demonstrate that measure optimization returns the correct CP decomposition under an incoherence condition on the rank-one factors. To address the computational challenge, we present a hierarchy of semidefinite programs based on sums-of-squares relaxations of the measure optimization problem. By showing that the constructed dual polynomial is a sum-of-squares modulo the sphere, we prove that the smallest SDP in the relaxation hierarchy is exact and the decomposition can be extracted from the solution under the same incoherence condition. One implication is that the tensor nuclear norm can be computed exactly using the smallest SDP as long as the rank-one factors of the tensor are incoherent. Numerical experiments are conducted to test the performance of the moment approach.

1. Introduction

Tensors provide a compact and natural representation for high-dimensional, multi-view datasets encountered in fields such as communication, signal processing, large-scale data analysis, and computational neuroscience, to name a few. In many data analysis tasks, tensor-based approaches outperform matrix-based ones due to the ability to identify non-orthogonal components, a property derived from having access to higher order moments (Landsberg, 2009). In this work, we investigate the problem of decomposing a tensor into a linear combination of a small number

of rank one tensors, also known as the CP decomposition or the PARAFAC decomposition. Such low-rank tensor decomposition extends the idea of singular value decomposition for matrices and finds numerous applications in data analysis (Papalexakis et al., 2013; Anandkumar et al., 2013; 2012; Cichocki et al., 2014; Comon, 2009; Kolda & Bader, 2009; Lim & Comon, 2010).

We approach tensor decomposition from the point of view of measure estimation from moments. To illustrate the idea, consider determining the CP decomposition of a third order symmetric tensor $A = \sum_{p=1}^r \lambda_p x^p \otimes x^p \otimes x^p$, which can be viewed as estimating a discrete measure $\mu^* = \sum_{p=1}^r \lambda_p \delta(x - x^p)$ supported on the unit sphere from its 3rd order moments $A_{ijl} = \int_{\mathbb{S}^{n-1}} x_i x_j x_l d\mu^*$. This formulation offers several advantages. First, it provides a natural way to enforce a low-rank decomposition by minimizing an infinite-dimensional generalization of the ℓ_1 norm, the total variation norm of the measure. Second, the optimal value of the total variation norm minimization problem, which is a convex problem in the space of measures, defines a norm for tensors. This norm, termed as the tensor nuclear norm, is an instance of atomic norms, which, as argued by the authors of (Chandrasekaran et al., 2012), is the best possible convex proxy for recovering simple models. Just like the matrix nuclear norm, the tensor nuclear norm can be used to enforce low-rankness in tensor completion, robust tensor principal component analysis, and stable tensor recovery. Finally, finite computational schemes developed for atomic tensor decomposition can be readily modified to accomplish these more complex tensor tasks.

The theoretical analysis of atomic tensor decomposition is fundamental in understanding the regularization and estimation power of the tensor nuclear norm in solving other tensor problems. For instance, it tells us what types of rank-one tensor combinations are identifiable given full, noise-free data. Moreover, the dual polynomial constructed to certify a particular decomposition is useful in investigating the performance of tensor nuclear norm minimization for data corrupted by missing observations, noise, and outliers. This work parallels similar measure estimation ideas as applied to line spectral estimation in signal processing

to achieve super-resolution, for which the dual polynomial constructed in (Candès & Fernandez-Granda, 2014) was later utilized to analyze the ability of frequency estimation from incomplete, noisy, and grossly corrupted data (Tang et al., 2014b;a; 2013; Chi & Chen, 2014). We expect that the tensor decomposition results will find similar uses in the corresponding tensor tasks.

Our contributions in this work are three-fold. Firstly, we formulate atomic tensor decomposition as a moment problem and apply the Lasserre sum-of-squares (SOS) relaxation hierarchy to obtain a series semidefinite programs (SDPs) to approximately solve the moment problem. Secondly, we explicitly construct a dual polynomial to certify that a decomposition with *incoherent* components $\{x^p, p = 1, \dots, r\}$ is the unique atomic tensor decomposition. The incoherence condition requires that the matrix formed by the vectors $\{x^p\}$ is well-conditioned. Lastly, by showing that the constructed dual polynomial is sum-of-squares modulo the sphere, we establish that the *smallest* SDP in the Lasserre hierarchy *exactly* solves the atomic decomposition under the *same* incoherent assumption. Such a result is different from existing approximation results for the Lasserre hierarchy, where there is no guarantee on the size of the SDP at which exact relaxation occurs (Nie, 2014). The effectiveness of the lowest order relaxations is of crucial importance for computation, as the Lasserre hierarchy is considered impractical due to the rapid increase of the sizes of SDPs in the hierarchy.

2. Connections to Prior Art

CP tensor decomposition is a classical tensor problem that has been studied by many authors (cf. (Comon, 2009; Kolda & Bader, 2009)). Most tensor decomposition approaches are based on alternating minimization, which typically do not offer global convergence guarantees (Bro, 1997; Harshman, 1970; Kolda & Bader, 2009; Papalexakis et al., 2013; Comon et al., 2009). However, recent work that combines alternating minimization and power iteration has yielded guaranteed tensor decomposition in a probabilistic setting where the algorithm is randomized (Anandkumar et al., 2013; 2014). In contrast, the theoretical guarantee of our moment approach is deterministic, which is more natural since there is no randomness in the problem formulation. Furthermore, our approach is flexible and capable of incorporating a variety of sources of uncertainty in the framework; this includes noise, missing data, partial measurements, and gross outliers.

Another closely related line of work is matrix completion and tensor completion. Low-rank matrix completion and recovery based on the idea of nuclear norm minimization has received a great deal of attention in recent years (Candès & Recht, 2009; Recht et al., 2010; Recht, 2011).

A direct generalization of this approach to tensors would be using tensor nuclear norm to perform low-rank tensor completion and recovery. However, this approach was not pursued due to the NP-hardness of computing the tensor nuclear norm (Hillar & Lim, 2013). The mainstream tensor completion approaches are based on various forms of matricization and application of matrix completion to the flattened tensor (Gandy et al., 2011; Liu et al., 2013; Mu et al., 2013; Shah et al., 2015). Alternating minimization can also be applied to tensor completion and recovery with performance guarantees established in recent work (Huang et al., 2014). Most matricization nor alternating minimization approaches do not yield optimal bounds on the number of measurements needed for tensor completion. One exception is (Shah et al., 2015), which uses a special class of separable sampling schemes.

In contrast, we expect that the atomic norm, when specialized to tensors, will achieve the information theoretical limit for tensor completion as it does for compressive sensing, matrix completion (Recht, 2011), and line spectral estimation with missing data (Tang et al., 2013). Given a set of atoms, the atomic norm is an abstraction of ℓ_1 -type regularization that favors simple models. Using the notion of descent cones, the authors of (Chandrasekaran et al., 2012) argued that the atomic norm is the best possible convex proxy for recovering simple models. Particularly, atomic norms were shown in many problems beyond compressive sensing and matrix completion to be able to recover simple models from minimal number of linear measurements. For example, when specialized to the atomic set formed by complex exponentials, the atomic norm can recover signals having sparse representations in the continuous frequency domain with the number of measurements approaching the information theoretic limit without noise (Tang et al., 2013), as well as achieving near minimax denoising performance (Tang et al., 2014a). Continuous frequency estimation using the atomic norm is also an instance of measure estimation from (trigonometric) moments.

The computational foundation of our approach is SOS relaxation, particularly the Lasserre hierarchy for moment problems (Parrilo, 2000; Lasserre, 2001). After more than a decade’s developments, SOS relaxations have produced a large body of literature (cf. (Blekherman et al., 2013), (Lasserre, 2009) and references therein). The Lasserre hierarchy provides a series of SDPs that can approximate moment problems with increasing tightness (Lasserre, 2001; Parrilo, 2000; Lasserre, 2008). Indeed, it has been shown that as one moves up the hierarchy, the solutions of the SDP relaxations converge to the infinite-dimensional measure optimization (Nie, 2014). In many cases, finite convergence is also possible, though it is typically hard to determine the sizes of those exact relaxations (Nie, 2014). We show that for the tensor decomposition problem, exact re-

laxation occurs for the smallest SDP in the hierarchy under certain incoherence conditions. Combining with the necessary condition in Theorem 2, we can roughly say that when the atomic tensor decomposition is solvable by the total variation norm minimization, it is also solvable by a small SDP; when the lowest order SDP relaxation does not work, the original infinite-dimensional measure optimization is also unlikely to work. A very recent piece of work (Barak & Moitra, 2015) utilizes the SOS hierarchy for tensor completion in the presence of noise, and studies both upper and lower bounds. While related, this work focuses more on the noisy performance case. We are also able to avoid the negative results of (Barak & Moitra, 2015) by assuming that the target tensor’s factors are incoherent.

3. Model and Algorithm

3.1. Model for tensor decomposition

We focus on third order, symmetric tensors in this work. Given such a tensor $A = [A_{ijkl}]_{i,j,l=1}^n \in S^3(\mathbb{R}^n)$, we are interested in decompositions of the form

$$A = \sum_{p=1}^r \lambda_p x^p \otimes x^p \otimes x^p \quad (1)$$

where $\|x^p\| = 1$ and $\lambda_p > 0$. The decomposition expressing a tensor as the sum of a finite number of rank-one tensors is called the CP decomposition (Canonical Polyadic Decomposition), which also goes by the name of CANDECOMP (Canonical Decomposition) (Carroll & Chang, 1970) and PARAFAC (Parallel Factors Decomposition) (Harshman, 1970). The positive coefficient assumption does not reduce the generality of the model since the sign of λ_p can be absorbed into the vector x^p . The smallest r that allows such a decomposition is called the symmetric rank of A , denoted by $\text{srnk}(A)$. A decomposition with $\text{srnk}(A)$ terms is always possible, though like many other tensor problems, determining the symmetric rank of a general 3rd-order, symmetric tensor is NP-hard (Hillar & Lim, 2013).

Denote the unit sphere of \mathbb{R}^n as \mathbb{S}^{n-1} , and the set of (non-negative) Borel measures on \mathbb{S}^{n-1} as $\mathcal{M}_+(\mathbb{S}^{n-1})$. We write the CP decomposition in (1) as

$$A = \int_{\mathbb{S}^{n-1}} x \otimes x \otimes x d\mu^* \quad (2)$$

where the decomposition measure $\mu^* = \sum_{p=1}^r \lambda_p \delta(x - x^p) \in \mathcal{M}_+(\mathbb{S}^{n-1})$. Hereafter, we use a superscript \star to indicate that the measure is the “true”, unknown decomposition measure to be identified from the tensor A . Since the entries of A are 3rd order moments of the measure μ^* , tensor decomposition is an instance of measure estimation from moments. Model (2) is more general than (1) in the sense that it allows decompositions involving infinite num-

ber of rank-one tensors. However, in most cases the decompositions of interest involve finite terms. In particular, we restate the problem of determining $\text{srnk}(A)$ as

$$\underset{\mu \in \mathcal{M}_+(\mathbb{S}^{n-1})}{\text{minimize}} \|\mu\|_0 \text{ subject to } A = \int_{\mathbb{S}^{n-1}} x \otimes x \otimes x d\mu \quad (3)$$

where $\|\mu\|_0$ is the support size of μ . This is a generalization of the ℓ_0 “norm” minimization problem to the infinite-dimensional measure space.

Following the idea of using the ℓ_1 norm as a convex proxy for the ℓ_0 “norm” and recognizing “ $\|\mu\|_{\ell_1} = \mu(\mathbb{S}^{n-1})$ ”, we formulate symmetric tensor decomposition as the following optimization

$$\begin{aligned} & \underset{\mu \in \mathcal{M}_+(\mathbb{S}^{n-1})}{\text{minimize}} \mu(\mathbb{S}^{n-1}) \\ & \text{subject to } A = \int_{\mathbb{S}^{n-1}} x \otimes x \otimes x d\mu \end{aligned} \quad (4)$$

Note that the total mass $\mu(\mathbb{S}^{n-1})$ is the restriction of the total variation norm to the set of (non-negative) Borel measures. For any third order symmetric tensor A , the optimal value of (4) defines the tensor nuclear norm $\|A\|_*$, which is a special case of the more general atomic norms. A consequence of Caratheodory’s theorem concerning convex hulls (Barvinok, 2002) is that there always exists an optimal solution with finite support. We call a decomposition corresponding to an optimal, finite measure solving (4) an *atomic tensor decomposition*.

The optimization (4) is an instance of the problem of moments (Lasserre, 2008), whose dual is

$$\begin{aligned} & \underset{Q \in S^3(\mathbb{R}^n)}{\text{maximize}} \langle Q, A \rangle \\ & \text{subject to } \langle Q, x \otimes x \otimes x \rangle \leq 1, \forall x \in \mathbb{S}^{n-1}. \end{aligned} \quad (5)$$

We have used $\langle A, B \rangle = \sum_{i,j,l} A_{ijl} B_{ijl}$ to denote the inner product of two 3rd order tensors. The homogeneous polynomial $q(x) := \langle Q, x \otimes x \otimes x \rangle = \sum_{i,j,k} Q_{ijk} x_i x_j x_k$ corresponding to a dual feasible solution is called a dual polynomial. We will see that the dual polynomial associated with the optimal dual solution can be used to certify the optimality of a particular decomposition.

3.2. Moment Relaxation

The tensor decomposition problem (4) is a special truncated moment problem (Nie, 2012), where we observe only third order moments of a measure μ^* supported on the unit sphere. Therefore, we can apply the Lasserre SDP hierarchy (Lasserre, 2001) to approximate the infinite dimensional linear program (4). We first introduce a few notations in order to describe the SDP hierarchy. We use $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{N}^n$ to denote a multi-integer index. The notation x^α represents the monomial $x_1^{\alpha_1} x_2^{\alpha_2} \dots x_n^{\alpha_n}$. The size of α , $|\alpha| = \sum \alpha_i$, is the degree of x^α . The set

$\mathbb{N}_k^n = \{\alpha : |\alpha| \leq k\} \subset \mathbb{N}^n$ consists of indices with sizes less than or equal to k . The notation $\mathbb{R}^{\mathbb{N}_k^n}$ ($\mathbb{R}^{\mathbb{N}_k^n \times \mathbb{N}_k^n}$, resp.) represents the set of real vectors (matrices, resp.) whose entries are indexed by elements in \mathbb{N}_k^n ($\mathbb{N}_k^n \times \mathbb{N}_k^n$, resp.).

For $k = 2, 3, 4, \dots$ and a vector $z \in \mathbb{R}^{\mathbb{N}_{2k}^n}$, we use the matrix $M_k(z) \in \mathbb{R}^{\mathbb{N}_k^n \times \mathbb{N}_k^n}$ to denote the moment matrix associated with the vector $z \in \mathbb{R}^{\mathbb{N}_{2k}^n}$, whose (α, β) th entry is $z_{\alpha+\beta}$ for $\alpha, \beta \in \mathbb{N}_k^n$. The notation $L_{k-1}(z) \in \mathbb{R}^{\mathbb{N}_{k-1}^n \times \mathbb{N}_{k-1}^n}$ is reserved for the localizing matrix of the polynomial $p(x) := \|x\|_2^2 - 1$, whose (α, β) th entry is

$$[L_{k-1}(z)]_{\alpha, \beta} = \sum_{|\gamma| \leq 2} p_\gamma z_{\alpha+\beta+\gamma}, \alpha, \beta \in \mathbb{N}_{k-1}^n \quad (6)$$

with p_γ the coefficient for the monomial x^γ in $p(x)$.

Each measure μ supported on \mathbb{S}^{n-1} is associated with an infinite sequence $\bar{y} \in \mathbb{R}^{\mathbb{N}^n}$, called the moment sequence, via $\bar{y}_\alpha := \int_{\mathbb{S}^{n-1}} x^\alpha d\mu, \alpha \in \mathbb{N}^n$. When $\alpha = (0, 0, \dots, 0)$, we use $\bar{y}_0 = \int_{\mathbb{S}^{n-1}} 1 d\mu = \mu(\mathbb{S}^{n-1})$ to denote the total mass of the measure μ . Denote by $\mathfrak{M}(\mathbb{S}^{n-1}) \subset \mathbb{R}^{\mathbb{N}^n}$ the set of all such moment sequences. Instead of optimizing with respect to a measure μ in (4), we can equivalently optimize with respect to a moment sequence:

$$\text{minimize } \bar{y}_0 \text{ subject to } \bar{y}_\alpha = A_{ijk} \text{ if } x^\alpha = x_i x_j x_k \quad (7)$$

The constraint that $\bar{y} \in \mathfrak{M}(\mathbb{S}^{n-1})$ involves an infinite sequence. To obtain a finite optimization problem, we relax (7) by replacing \bar{y} with its $2k$ -truncation $y \in \mathbb{R}^{\mathbb{N}_{2k}^n}$ and replacing the constraint $\bar{y} \in \mathfrak{M}(\mathbb{S}^{n-1})$ with the following easy-to-enforce conditions for y to be the finite truncation of a moment sequence (Lasserre, 2009):

$$M_k(y) \succcurlyeq 0, L_{k-1}(y) = 0. \quad (8)$$

The validity of the linear matrix inequality constraint and equality constraint is standard and uses the fact that μ is a Borel (hence nonnegative) measure on \mathbb{S}^{n-1} and $p(x) = \|x\|_2^2 - 1 \equiv 0$ on \mathbb{S}^{n-1} .

Hence, we obtain a finite-dimensional relaxation for (7):

$$\text{minimize } y_0 \text{ subject to } y_\alpha = A_{ijl} \text{ if } x^\alpha = x_i x_j x_l$$

$$y \in \mathbb{R}^{\mathbb{N}_{2k}^n} \quad M_k(y) \succcurlyeq 0, L_{k-1}(y) = 0. \quad (9)$$

Denote by $\|A\|_{k,*}$ the optimal value of (9). One can verify that $\|\cdot\|_{k,*}$ indeed defines a norm in the space of symmetric tensors. Clearly $\|A\|_{k,2}$ is smaller than $\|A\|_*$ for all symmetric tensors A and increasing k (i.e., using longer truncation) allows us to get better approximations. In a more general setting, it has been shown that the optimal value of (9) converges to that of (4) as $k \rightarrow \infty$ even in finite steps (Nie, 2014). Furthermore, if the moment matrix $M_k(\hat{y})$ associated with the optimal solution \hat{y} of (9) satisfies the flat extension condition, i.e., $\text{rank}(M_k(\hat{y})) = \text{rank}(M_{k-1}(\hat{y}))$, then we can apply an algebraic procedure to recover the

measure $\hat{\mu}$ from the moment matrix (Curto & Fialkow, 1996; Henrion & Lasserre, 2005). Our goal is to show that under reasonable conditions on the true decomposition of the measure μ^* that generates observations in A , the smallest relaxation with $k = 2$ is exact, i.e., $\|A\|_{2,*} = \|A\|_*$, and is sufficient for the recovery of μ^* .

By following a standard procedure of deriving the Lagrange dual, we get the following dual problem of (9):

$$\text{maximize } \langle Q, A \rangle$$

$$Q \in S^3(\mathbb{R}^n), H, G$$

$$\text{subject to } e_0 - 1_\Omega(\text{vec}(Q)) = M_k^*(H) + L_{k-1}^*(G)$$

$$H \succcurlyeq 0 \quad (10)$$

Here $*$ represents the adjoint operator; $e_0 \in \mathbb{R}^{\mathbb{N}_{2k}^n}$ denotes the first canonical basis vector; and the operation $\text{vec}(Q)$ takes the unique entries in the symmetric tensor Q to form a vector, which is then embedded by 1_Ω into the third order moment vector space $\mathbb{R}^{\mathbb{N}_3^n}$.

We show that (9) is an SOS relaxation by rewriting its dual (10) as an SOS optimization. For this purpose, we denote the vector consisting of all monomials of x of degrees up to k by $\nu_k(x)$, also known as the Veronese map. For example, when $k = 2$, $\nu_2(x)$ has the following form:

$$\nu_2(x) = [1 \quad x_1 \quad \dots \quad x_n \quad x_1^2 \quad x_1 x_2 \quad \dots \quad x_n^2]^T \quad (11)$$

Define two polynomials $\sigma(x) := \nu_k(x)' H \nu_k(x)$ and $s(x) := \nu_{k-1}(x)' G \nu_{k-1}(x)$ for feasible solutions H and G of (10). Since $H \succcurlyeq 0$, the polynomial $\sigma(x)$ is the Gram matrix representation of an SOS polynomial and $s(x)$ is an arbitrary polynomial of degree $2k - 2$. We now rewrite the optimization (10) as an SOS optimization:

$$\text{maximize } \langle Q, A \rangle$$

$$Q \in S^3(\mathbb{R}^n)$$

$$\text{subject to } 1 - q(x) = s(x)(\|x\|_2^2 - 1) + \sigma(x)$$

$$\deg(s(x)) \leq 2k - 2$$

$$\sigma(x) \text{ is an SOS with } \deg(\sigma(x)) \leq 2k \quad (12)$$

where $q(x) = \langle Q, x \otimes x \otimes x \rangle$ is the dual polynomial defined before. Compared with the dual polynomial in (5), the one here $q(x) = 1 - \sigma(x) - s(x)(\|x\|_2^2 - 1)$ is more structured. We call $1 - q(x)$ an SOS modulo the sphere.

4. Main Results

The main theorem of this work relies on the construction of dual polynomials that certify the optimality of the decomposition measure μ^* . Due to space limitation, the detailed construction of the dual polynomials are deferred to the supplemental materials. The constructed dual polynomials are also essential to the development of noise performance and tensor completion results using the moment approach. We record the following proposition, which forms the basis

of the dual polynomial proof technique.

Proposition 1. Suppose $\text{supp}(\mu^*) = \{x^p, p = 1, \dots, r\}$ is such that $\{x^p \otimes x^p \otimes x^p, p = 1, \dots, r\}$ forms a linearly independent set.

1. If there exists a $Q \in S^3(\mathbb{R}^n)$ such that the associated dual polynomial $q(x)$ satisfies

$$q(x^p) = 1, p = 1, \dots, r \quad (13)$$

$$q(x) < 1, x \neq x^p, \forall p, \quad (14)$$

then $\mu^* = \sum_{p=1}^r \lambda_p \delta(x - x^p)$ is the unique solution of the moment problem (4).

2. In addition to part 1, the dual polynomial $q(x)$ also has the form $1 - \sigma(x) - s(x)(\|x\|_2^2 - 1)$, where $\sigma(x)$ is an SOS with $\deg(\sigma(x)) \leq 2k$, and $\deg(s(x)) \leq 2(k-1)$, then the optimization (9) is an exact relaxation of (4), *i.e.*, $\|A\|_{k,*} = \|A\|_*$. Furthermore, y^* , the $2k$ -truncation of the moment sequence for μ^* is an optimal solution to (9).

3. Suppose $\{x^p, p = 1, \dots, r\}$ are linearly independent. (So $r \leq n$.) In addition to the conditions in parts 1 and 2, if the Gram matrix H for the SOS $\sigma(x)$ in part 2 has rank $|\mathbb{N}_k^n| - r$, then y^* , the $2k$ -truncation of the moment sequence for μ^* , is the unique solution to (9) and we can extract the measure μ^* from the moment matrix $M_k(y^*)$.

A dual polynomial satisfying the interpolation and boundedness conditions (13) and (14) is used frequently as the starting point to derive several atomic decomposition and super-resolution results (Candès & Fernandez-Granda, 2014; Tang & Recht; Bendory et al., 2014b;a; Heckel et al., 2014). The second part of Proposition 1, which additionally requires the polynomial be SOS modulo the sphere to certify the exact relaxation of the SDP (9), is a contribution of this work. Part 3 is a consequence of the flat extension condition (Curto & Fialkow, 1998; 1996). We remark that there is a version of part 3 that allows $r > n$ under additional assumptions, which we did not present here. Constructing a structured dual polynomial satisfying conditions in parts 2 and 3 allows us to identify the class of polynomial-time solvable instances of the tensor decomposition problem, which are NP hard in the worst case.

We are now ready to state our major theorem:

Theorem 1. For a symmetric tensor $A = \sum_{p=1}^r \lambda_p x^p \otimes x^p \otimes x^p$, if the vectors $\{x^p\}$ are incoherent, that is, the matrix $X = [x^1, x^2, \dots, x^r]$ satisfies

$$\|X^T X - I_r\| \leq 0.0016, \quad (15)$$

then there exists a dual symmetric tensor Q such that the dual polynomial $q(x) = \langle Q, x \otimes x \otimes x \rangle$ satisfies the conditions in all three parts of Proposition 1 with $k = 2$. Thus, $A = \sum_{p=1}^r \lambda_p x^p \otimes x^p \otimes x^p$ is the unique decomposition that achieves both the tensor nuclear norm $\|A\|_*$ and its relaxation $\|A\|_{2,*}$. Furthermore, this unique decomposition can be also found by solving (9) with $k = 2$.

A few remarks follow. The constant 0.0016 in the incoherent condition (15) has not been optimized.

The condition (15) requires $r \leq n$, which seems weak considering that the generic rank of a 3rd order symmetric tensor is at least $\frac{(n+2)(n+1)}{6}$ for all n except $n = 5$ (Comon et al., 2008). Furthermore, the Kruskal's sufficient condition states that a decomposition $A = \sum_{p=1}^r \lambda_p x^p \otimes x^p \otimes x^p$ is unique as long as $r \leq \frac{3k_X - 2}{2}$ where k_X is the Kruskal rank, or the maximum value of k such that any k columns of the matrix $X = [x^1, \dots, x^r]$ are linearly independent (Landsberg, 2009). Since $k_X \leq n$, the Kruskal rank condition is valid for r as large as $\frac{3n-2}{2}$.

There are two reasons for the requirement of $r \leq n$. The first one is technical: we used a perturbation analysis of the orthogonal symmetric tensor decomposition, which prevents $r > n$ in the first place. The second reason is due to the use of $k = 2$ in the relaxation (9) and is more fundamental. In order to extract the decomposition, we apply the flat extension condition $M_1(y) = M_2(y)$ and the procedure developed in (Henrion & Lasserre, 2005). Since the size of $M_1(y)$ is $n + 1$, there is no way to identify more than $n + 1$ components from the moment matrix $M_2(y)$. If the goal is extract the decomposition from the moment matrix, as addressed in this paper, we will need to increase the relaxation to $k \geq 3$ to recover decompositions with more than $n + 1$ components. However, if the goal is denoising or tensor completion, it is still possible to achieve optimal noise performance and exact completion using $k = 2$ even if $r > n + 1$. Indeed, numerical experiments in Section 6.2 show that the smallest SDP of (9) can recover all moments up to order 4 correctly for r as large as $2n$.

To complement the sufficient condition in Theorem 1, we cite a theorem of (Tang, 2015) which demonstrates that a separation or incoherence condition on $\{x^p\}$ is necessary.

Theorem 2. (Tang, 2015) Consider a set of vectors $S = \{x^p, p = 1, \dots, r\} \subset \mathbb{S}^{n-1}$. If any signed measure supported on S is the unique solution to the optimization

$$\underset{\mu \in \mathcal{M}(\mathbb{S}^{n-1})}{\text{minimize}} \|\mu\|_{\text{TV}} \text{ subject to } A = \int_{\mathbb{S}^{n-1}} x^{m \otimes} d\mu \quad (16)$$

then the maximal incoherence of points in S satisfies

$$\max_{i \neq j} (|\langle x^i, x^j \rangle|) \leq \cos(2/m). \quad (17)$$

Here $\mathcal{M}(\mathbb{S}^{n-1})$ is the set of all signed measures on \mathbb{S}^{n-1} and $\|\cdot\|_{\text{TV}}$ denotes the total variation norm of a measure.

The incoherence condition (17) is a separation condition on points on \mathbb{S}^{n-1} as it is equivalent to that the angle between any two points x^i and x^j is greater than $2/m$. The upper bound in (17) further confirms that knowledge of higher moments reduces the incoherence requirement. Note that when m is odd, we can again focus on Borel (non-negative)

measures supported on $S = \{\pm x^p, p = 1, \dots, r\}$, and the total variation norm $\|\mu\|_{\text{TV}}$ can be replaced by the total mass $\mu(\mathbb{S}^{n-1})$. We also observe that the incoherence condition (15) in Theorem 1 for 3rd order symmetric tensor implies $\max_{i \neq j} (|\langle x_i, x_j \rangle|) \leq 0.0016 < \cos(2/3) \approx 0.7859$, which is stronger than the necessary condition (17).

5. Extensions

5.1. Tensor completion and denoising

Since the optimal value of (4) defines the tensor nuclear norm $\|\cdot\|_*$, the results developed for tensor decomposition will form the foundation for tensor completion and stable low-rank tensor recovery. Similar to its matrix counterpart, the tensor nuclear norm favors low-rank solutions when the observations are corrupted by noise, missing data, and outliers. For example, when a low-rank tensor A^* is partially observed on an index set Ω , we can fill in the missing entries by solving a tensor nuclear norm minimization problem (Jain & Oh, 2014; Acar et al., 2011; Yuan & Zhang, 2014; Huang et al., 2014; Gandy et al., 2011):

$$\underset{A}{\text{minimize}} \|A\|_* \text{ subject to } A_\Omega = A_\Omega^*. \quad (18)$$

This line of thinking was previously considered infeasible due to the intractability of the tensor nuclear norm. However, we can use the relaxed norm $\|\cdot\|_{k,*}$ to approximate (18):

$$\underset{A}{\text{minimize}} \|A\|_{k,*} \text{ subject to } A_\Omega = A_\Omega^*. \quad (19)$$

which is equivalent to the SDP:

$$\begin{aligned} & \underset{y \in \mathbb{R}^{\binom{n}{2k}}}{\text{minimize}} y_0 \\ & \text{subject to } y_\alpha = A_{ijl} \text{ when } x^\alpha = x_i x_j x_l \text{ and } (i, j, l) \in \Omega \\ & \quad M_k(y) \succcurlyeq 0, L_{k-1}(y) = 0. \end{aligned} \quad (20)$$

Building on the dual polynomial of Theorem 1, we expect to show that $\|\cdot\|_{2,*}$ can be used to perform completion with a minimal number of tensor measurements, given that the tensor factors are incoherent.

Gaussian-type noise, which is unavoidable in practical scenarios, can also be handled using the tensor nuclear norm:

$$\underset{A}{\text{minimize}} \frac{1}{2} \|A - B\|_2^2 + \gamma \|A\|_*, \quad (21)$$

where B is the observed noisy entries of the tensor and γ is a regularization parameter. Replacing $\|\cdot\|_*$ with $\|\cdot\|_{k,*}$ gives rise to a hierarchy of SDP relaxations for (21):

$$\begin{aligned} & \underset{y \in \mathbb{R}^{\binom{n}{2k}}}{\text{minimize}} \frac{1}{2} \|A - B\|_2^2 + \gamma y_0 \\ & \text{subject to } y_\alpha = A_{ijl} \text{ when } x^\alpha = x_i x_j x_l \\ & \quad M_k(y) \succcurlyeq 0, L_{k-1}(y) = 0. \end{aligned} \quad (22)$$

We conducted numerical experiments to demonstrate the performance of tensor completion and denoising using the smallest SDP relaxations in (20) and (22).

5.2. Non-symmetric and high-order tensors

We briefly discuss extensions to non-symmetric and high-order tensor problems. Consider decomposing a non-symmetric tensor $A = [A_{ijk}] \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ into the form $A = \sum_{p=1}^r \lambda_p u^p \otimes v^p \otimes w^p$, where $\|u^p\| = \|v^p\| = \|w^p\| = 1$ and $\lambda_p > 0$. Similar to (4), we formulate the non-symmetric tensor decomposition again as estimating a measure μ supported on $K = \mathbb{S}^{n_1} \times \mathbb{S}^{n_2} \times \mathbb{S}^{n_3}$:

$$\underset{\mu \in \mathcal{M}_+(K)}{\text{minimize}} \mu(K) \text{ subject to } A = \int_K u \otimes v \otimes w d\mu. \quad (23)$$

Optimization (23) admits a similar SDP relaxation hierarchy for $k = 2, 3, \dots$:

$$\begin{aligned} & \underset{y \in \mathbb{R}^{\binom{n}{2k}}}{\text{minimize}} y_0 \\ & \text{subject to } y_\alpha = A_{ijl} \text{ when } \xi^\alpha = u_i v_j w_l \\ & \quad M_k(y) \succcurlyeq 0, L_{k-1}^{h_1}(y), L_{k-1}^{h_2}(y), L_{k-1}^{h_3}(y) = 0, \end{aligned} \quad (24)$$

where $\xi = (u, v, w)$, and $\{L_{k-1}^{h_i}\}$ are localizing matrices corresponding to the constraints $h_1(\xi) = \|u\|_2^2 - 1 = 0$, $h_2(\xi) = \|v\|_2^2 - 1 = 0$ and $h_3(\xi) = \|w\|_2^2 - 1 = 0$. The SDPs in (24) can be modified to solve tensor completion and denoising problems.

The measure formulation extends easily to higher-order tensors. For the SDP relaxation hierarchy, we just need to fill in the moment vector with the observed, high-order moments, and add more constraints corresponding to the constraints defining the measure domain K . However, theoretical treatment might be more challenging, especially if we would like to allow the rank r to go beyond the individual tensor dimensions.

6. Numerical Experiments

We performed a series of experiments to illustrate the performance of the SDP relaxations (9) in solving the tensor decomposition and other related problems. All the SDPs are solved using the CVX package.

6.1. Phase transitions with full data

Figure 1 shows the phase transitions for the success rate of the SDP relaxation (9) with $k = 2$ when we vary the rank r , the incoherence $\Delta = \max_{i \neq j} |\langle x^i, x^j \rangle|$, and the dimension n . The purpose is to figure out the critical incoherence value. In preparing the upper plot in Figure 1, we took $n = 10$, $r \in \{2, 4, \dots, 20\}$, and $\Delta \in \{0.38, 0.39, \dots, 0.52\}$.

We choose the maximal incoherence Δ instead of the quantity in condition (15) because in the experiments the rank r goes beyond n , in which case condition (15) is always violated. To compute the success rate, we produced $T = 10$ instances for each (r, Δ) configuration. We used the acceptance-rejection method to generate an instance with r vectors such that $\max_{i \neq j} |\langle x^i, x^j \rangle| \leq \Delta$. This method becomes inefficient when $\Delta < 0.38$, forcing us to test Δ s in the chosen range. After being passed through the SDP (9), an instance is declared success if the difference between the recovered moment vector and the true moment vector has an ℓ_2 norm less than 10^{-4} . Again, we choose this success criterion instead of correct identification of the decomposition because the rank r goes beyond n , in which case we can not identify the decomposition from the moment matrix. It is easy to see that when $r \leq n$ and the 4th order moment matrix is recovered correctly, the flat extension condition is satisfied and the decomposition can be extracted from the moment matrix.

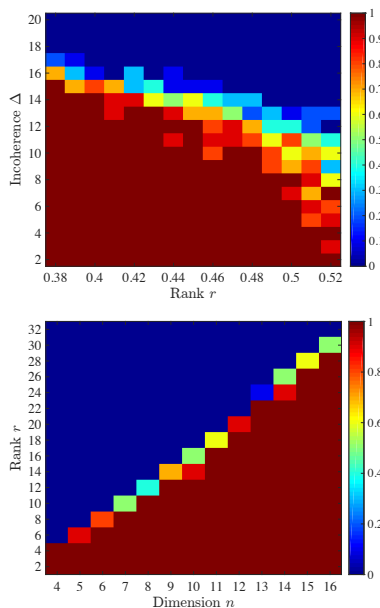


Figure 1. Color coded success rates of the lowest order SDP in recovering the 4th order moment vector: rank r vs incoherence Δ (Upper) for $n = 10$ and rank r vs dimension n for $\Delta = 0.38$.

We observe from the upper plot of Figure 1 that the incoherence condition can be relaxed for smaller tensors with smaller ranks. For rank $r \leq n = 10$, a critical separation of $\Delta = 0.45$ is sufficient for exact recovery. Though the figure does not show the transition for $\Delta < 0.38$ due to the difficulty of generating vectors maintaining such a small incoherence, by extrapolation we expect that when $\Delta \leq 0.38$, the relaxation (9) can recover instances with rank up to $r = 15 = \frac{3n}{2}$. We comment that the limitation to the range $\Delta \leq 0.38$ is due to the inefficiency of our re-

jection sampling methods to generate vectors with maximal incoherence smaller than 0.38. There are many vector configurations with a far smaller incoherence (Rankin, 1955), but we are not aware of an efficient algorithms to generate them (except for the orthogonal ones).

In the next experiment, we examine the phase transition when the dimension n and the rank r are varied while the incoherence Δ is fixed to 0.38. The purpose is to determine the critical rank r when the vectors $\{x^p\}$ are well-separated. We observe a clear phase transition, whose boundary is roughly $r = 2.1n - 6.4$.

6.2. Phase transition for completion

In this set of experiments, we test the power of the SDP relaxation (20) in performing symmetric tensor completion. In figure 2, we plot the success rates for tensors with *orthogonal* components when the number of observations, the rank r , and the dimension n are varied. To compute the success rate, the following procedure was repeated 10 times for each (m, r) or (m, n) configuration, where m is the number of measurements. A set of r random, orthonormal vectors $\{x^p\}$ together with a vector $\lambda \in \mathbb{R}^r$ following the uniform distribution on $[0, 1]$ were generated to produce the tensor $A = \sum_{p=1}^r \lambda_p x^p \otimes x^p \otimes x^p$. A uniform random subset of the tensor entries were sampled to form the observations. Since symmetric tensors have duplicated entries, we made sure only the unique entries were sampled and counted towards the measurements. The optimization (20) was then run to complete the tensor as well as estimating all the moments up to order 4. The optimization was successful if the ℓ_2 norm between the recovered 4th order moment vector and the true moment vector is less than 10^{-4} . We applied the same procedure to prepare the phase transition plots in Figure 3 except that the vectors $\{x^p\}$ are not orthogonal, but rather maintain an incoherence $\max_{i \neq j} |\langle x^i, x^j \rangle| \geq 0.38$.

For orthogonal tensor completions shown in Figure 2, we observe clear phase transitions for both the number of measurements versus the rank r , and versus the dimension n . Even though the degree of freedom for a dimension n , rank r , third-order symmetric tensor is rn , which is linear in both r and n , the boundaries in both plots of Figure 2 are curved. This phenomenon is seen in other completion tasks such as matrix completion (Candès & Recht, 2009) and compressed sensing off the grid (Tang et al., 2013). For non-orthogonal tensor completion, the phase transition boundaries are more blurred as seen from Figure 3. We believe this is because our selected value for the incoherence, 0.38, is still too large.

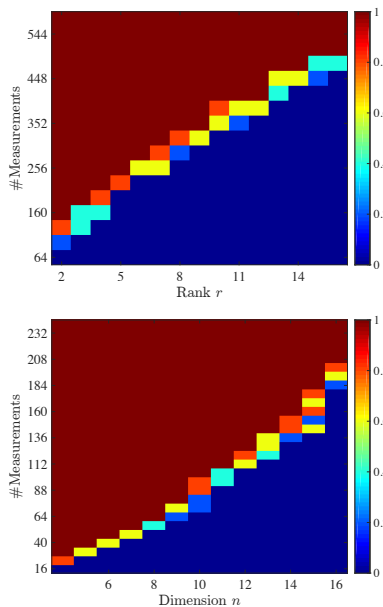


Figure 2. Color coded success rate of the lowest order SDP for *orthogonal* symmetric tensor completion: the number of measurements vs. rank r for fixed $n = 16$ (Upper), and the number of measurements vs. dimension n for fixed $r = 4$ (Lower).

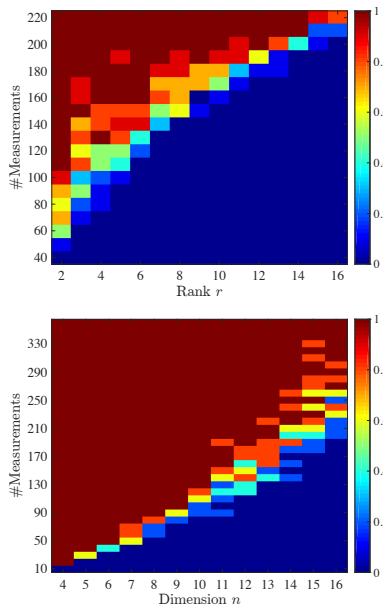


Figure 3. Color coded success rate of the lowest order SDP for non-orthogonal, symmetric tensor completion: the number of measurements vs. rank r for fixed $n = 10$ (Upper), and the number of measurements vs. dimension n for fixed $r = 4$ (Lower).

6.3. Noise robustness

In the last experiment, we show one example to demonstrate that the moment approach for tensor recovery is ro-

bust to Gaussian type noise. For $n = 5$, we generated a tensor with $r = 6$ random rank-one factors maintaining an incoherence less than 0.38. Gaussian noise of standard deviation σ equal to half the average magnitude of the tensor elements was added to all the unique entries of the tensor. We then ran the optimization (22) with $k = 2$ to perform denoising. The penalization parameter γ is set to equal σ . The noise-free and recovered 4th order moment vectors (except for the 0th order moments), and the observations are plotted in Figure 4. Note only 3rd order moments are observed while the algorithm returns all moments up to order 4. We chose to remove the 0th order moments because they are large and including them makes the plot less discernible. We see from Figure 4 that in addition to denoise the observed 3rd moments, which are entries of the tensor, the algorithm can also interpolate 0th to 2nd order moments and extrapolates the 4th order moments.

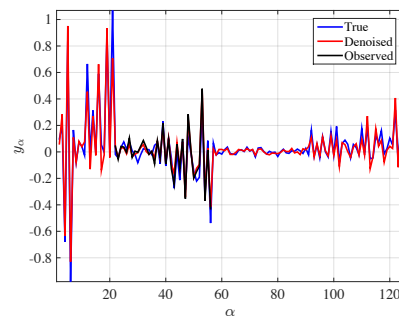


Figure 4. Tensor denoising using (22). Best viewed in color.

7. Conclusions

In this work, we formulated tensor decomposition as a measure estimation problem from observed moments, and used the total mass minimization to seek for a low-rank CP decomposition. We approximate this infinite-dimensional measure optimization using a hierarchy of SDPs. For third order symmetric tensors, by explicitly constructing an interpolation dual polynomial, we established that tensor decomposition is possible using the moment approach under an incoherence condition. Furthermore, by showing that the constructed dual polynomial is a sum-of-square modulo the sphere, we demonstrated that the smallest SDP in the relaxation hierarchy is exact, and the CP tensor decomposition can be identified from the recovered, truncated moment matrix. A complimentary resolution limit result was cited to show that certain incoherent condition was necessary. We discussed possible extensions to non-symmetric, and higher-order tensors, as well as generalizations to tensor completion and denoising. Numerical experiments were performed to test the power of the moment approach in tensor decomposition, completion, and denoising.

References

- Acar, Evrim, Dunlavy, Daniel M, Kolda, Tamara G, and Mørup, Morten. Scalable Tensor Factorizations for Incomplete Data. *Chemometrics and Intelligent Laboratory Systems*, 106(1):41–56, March 2011.
- Anandkumar, A, Anandkumar, Animashree, Ge, R, Ge, Rong, Janzamin, M, and Janzamin, Majid. Guaranteed Non-Orthogonal Tensor Decomposition via Alternating Rank-1 Updates. *arXiv.org*, 2014.
- Anandkumar, Anima, Ge, Rong, Hsu, Daniel, Kakade, Sham M, and Telgarsky, Matus. Tensor Decompositions for Learning Latent Variable Models. *arXiv.org*, October 2012.
- Anandkumar, Anima, Ge, Rong, Hsu, Daniel, and Kakade, Sham M. A Tensor Approach to Learning Mixed Membership Community Models. *arXiv.org*, February 2013.
- Barak, Boaz and Moitra, Ankur. Tensor prediction, rademacher complexity and random 3-xor. 01 2015. URL <http://arxiv.org/abs/1501.06521>.
- Barvinok, Alexander. *A Course in Convexity*. American Mathematical Soc., 2002.
- Bendory, Tamir, Dekel, Shai, and Feuer, Arie. Super-resolution on the Sphere using Convex Optimization. *arXiv.org*, December 2014a.
- Bendory, Tamir, Dekel, Shai, and Feuer, Arie. Robust Recovery of Stream of Pulses using Convex Optimization. *arXiv.org*, December 2014b.
- Blekherman, Grigoriy, Parrilo, Pablo A, and Thomas, Rekha R. *Semidefinite Optimization and Convex Algebraic Geometry*. SIAM, March 2013.
- Bro, Rasmus. PARAFAC. Tutorial and applications. *Chemometrics and Intelligent Laboratory Systems*, 38(2):149–171, October 1997.
- Candès, Emmanuel J and Fernandez-Granda, Carlos. Towards a Mathematical Theory of Super-resolution. *Communications on Pure and Applied Mathematics*, 67(6):906–956, June 2014.
- Candès, Emmanuel J and Recht, Benjamin. Exact Matrix Completion via Convex Optimization. *Foundations of Computational Mathematics*, 9(6):717–772, December 2009.
- Carroll, J Douglas and Chang, Jih-Jie. Analysis of individual differences in multidimensional scaling via an n-way generalization of “Eckart-Young” decomposition. *Psychometrika*, 35(3):283–319, 1970.
- Chandrasekaran, Venkat, Recht, Benjamin, Parrilo, Pablo A, and Willsky, Alan S. The Convex Geometry of Linear Inverse Problems. *Foundations of Computational Mathematics*, 12(6):805–849, 2012.
- Chi, Y and Chen, Y. Compressive Two-Dimensional Harmonic Retrieval via Atomic Norm Minimization. *Signal Processing, IEEE Transactions on*, (99):1–1042, 2014.
- Cichocki, A, Mandic, D, Caiafa, C, Phan, A H, Zhou, G, Zhao, Q, and De Lathauwer, L. Tensor Decompositions for Signal Processing Applications From Two-way to Multiway Component Analysis. *arXiv.org*, March 2014.
- Comon, P, Luciani, X, and de Almeida, A L F. Tensor Decompositions, Alternating Least Squares and Other Tales. *Journal of Chemometrics*, 23(7-8):393–405, July 2009.
- Comon, Pierre. Tensor Decompositions, State of the Art and Applications. *arXiv.org*, May 2009.
- Comon, Pierre, Golub, Gene, Lim, Lek-Heng, and Mourrain, Bernard. Symmetric Tensors and Symmetric Tensor Rank. *dx.doi.org*, 30(3):1254–1279, September 2008.
- Curto, R E and Fialkow, L A. *Solution of the Truncated Complex Moment Problem for Flat Data*, volume 568. Memoirs of the American Mathematical Society, Providence, RI, 1996.
- Curto, Raul E and Fialkow, Lawrence A. *Flat Extensions of Positive Moment Matrices-Recursively Generated Relations*, volume 648. Memoirs of the American Mathematical Society, Providence, RI, January 1998.
- De Lathauwer, L. Decompositions of a higher-order tensor in block terms-Part I: Lemmas for partitioned matrices. *SIAM Journal on Matrix Analysis and Applications*, 2008.
- Gandy, Silvia, Recht, Benjamin, and Yamada, Isao. Tensor Completion and Low-N-Rank Tensor Recovery via Convex Optimization. *Inverse Problems*, 27(2):025010, February 2011.
- Harshman, R A. Foundations of the PARAFAC procedure: Models and conditions for an “explanatory” multi-modal factor analysis. *UCLA Working Papers in Phonetics*, 1970.
- Heckel, Reinhard, Heckel, R, Morgenshtern, V I, Morgenshtern, Veniamin I, and Soltanolkotabi, Mahdi. Super-Resolution Radar. *arXiv.org*, 2014.
- Henrion, Didier and Lasserre, Jean-Bernard. Detecting Global Optimality and Extracting Solutions in GloptiPoly. In *Positive Polynomials in Control*, pp. 293–310. Springer Berlin Heidelberg, 2005.

- Hillar, Christopher J and Lim, Lek-Heng. Most Tensor Problems are NP-Hard. *Journal of the ACM (JACM)*, 60(6):45–39, November 2013.
- Huang, B, Jain, Prateek, Mu, C, Oh, Sewoong, Goldfarb, D, and Wright, J. Provable Tensor Factorization with Missing Data. *arXiv.org*, June 2014.
- Jain, Prateek and Oh, Sewoong. Provable Tensor Factorization with Missing Data. *arXiv.org*, June 2014.
- Kolda, T G and Bader, B W. Tensor decompositions and applications. *SIAM Review*, 2009.
- Landsberg, J M. Kruskal’s Theorem. *arXiv.org*, February 2009.
- Lasserre, J B. Global Optimization with Polynomials and the Problem of Moments. *SIAM Journal on Optimization*, 2001.
- Lasserre, Jean B. A Semidefinite Programming Approach to the Generalized Problem of Moments. *Mathematical Programming*, 112(1):65–92, March 2008.
- Lasserre, Jean-Bernard. *Moments, Positive Polynomials and Their Applications*. World Scientific, October 2009.
- Lim, Lek-Heng and Comon, Pierre. Multiarray Signal Processing: Tensor Decomposition Meets Compressed Sensing. *arXiv.org*, February 2010.
- Liu, Ji, Musialski, P, Wonka, P, and Ye, Jieping. Tensor Completion for Estimating Missing Values in Visual Data. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(1):208–220, 2013.
- Mu, Cun, Huang, Bo, Wright, John, and Goldfarb, Donald. Square Deal: Lower Bounds and Improved Relaxations for Tensor Recovery. *arXiv.org*, July 2013.
- Nie, Jiawang. The A-truncated K-Moment Problem. *arXiv.org*, October 2012.
- Nie, Jiawang. Optimality Conditions and Finite Convergence of Lasserre’s Hierarchy. *Mathematical Programming*, 146(1-2):97–121, August 2014.
- Papalexakis, E, Kang, U, Faloutsos, C, and Sidiropoulos, N. Large Scale Tensor Decompositions: Algorithmic Developments and Applications. *IEEE Data Engineering Bulletin - Special Issue on Social Media*, 2013.
- Parrilo, P A. *Structured Semidefinite Programs and Semialgebraic Geometry Methods in Robustness and Optimization*. Ph. D. thesis, 2000.
- Rankin, R A. The Closest Packing of Spherical Caps in n Dimensions. *Proceedings of the Glasgow Mathematical Association*, 2(03):139–144, July 1955.
- Recht, B. A Simpler Approach to Matrix Completion. *The Journal of Machine Learning Research*, 2011.
- Recht, B, Fazel, M, and Parrilo, P A. Guaranteed Minimum-Rank Solutions of Linear Matrix Equations via Nuclear Norm Minimization. *SIAM Review*, 2010.
- Shah, Parikshit, Rao, Nikhil, and Tang, Gongguo. Optimal low-rank tensor recovery from separable measurements: Four contractions suffice. 05 2015. URL <http://arxiv.org/abs/1505.04085>.
- Tang, Gongguo. Resolution Limits for Atomic Decompositions via Markov-Bernstein type Inequalities. In *International Conferenc on Sampling Theory and Applications*, 2015.
- Tang, Gongguo and Recht, Benjamin. Atomic decomposition of mixtures of translation-invariant signals. *Preprint*.
- Tang, Gongguo, Bhaskar, B N, Shah, P, and Recht, B. Compressed Sensing Off the Grid. *Information Theory, IEEE Transactions on*, 59(11):7465–7490, 2013.
- Tang, Gongguo, Bhaskar, B N, and Recht, B. Near Mini-max Line Spectral Estimation. *to appear in Information Theory, IEEE Transactions on*, 2014a.
- Tang, Gongguo, Shah, Parikshit, Bhaskar, Badri Narayan, and Recht, Benjamin. Robust Line Spectral Estimation. In *Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, 2014b.
- Yuan, Ming and Zhang, Cun-Hui. On Tensor Completion via Nuclear Norm Minimization. *arXiv.org*, May 2014.