Efficient algorithms for tensor scaling, quantum marginals, and moment polytopes

Peter Bürgisser
Technische Universität
Berlin, Germany
Email: pbuerg@math.tu-berlin.de

Cole Franks
Rutgers University
New Brunswick, USA
Email: wcf17@math.rutgers.edu

Ankit Garg
Microsoft Research
Bengaluru, India
Email: garga@microsoft.com

Rafael Oliveira
University of Toronto
Toronto, Canada
Email: rafael@cs.toronto.edu

Michael Walter
University of Amsterdam, QuSoft
Amsterdam, Netherlands
Email: m.walter@uva.nl

Avi Wigderson
Institute for Advanced Study
Princeton, USA
Email: avi@ias.edu

***Abstract*—We present a polynomial time algorithm to approximately *scale* tensors of any format to *arbitrary* prescribed marginals (whenever possible). This unifies and generalizes a sequence of past works on matrix, operator and tensor scaling. Our algorithm provides an efficient weak membership oracle for the associated *moment polytopes*, an important family of implicitly-defined convex polytopes with exponentially many facets and a wide range of applications. These include the entanglement polytopes from quantum information theory (in particular, we obtain an efficient solution to the notorious one-body quantum marginal problem) and the Kronecker polytopes from representation theory (which capture the asymptotic support of Kronecker coefficients). Our algorithm can be applied to succinct descriptions of the input tensor whenever the marginals can be efficiently computed, as in the important case of matrix product states or tensor-train decompositions, widely used in computational physics and numerical mathematics.**

Beyond these applications, the algorithm enriches the arsenal of numerical methods for classical problems in invariant theory that are significantly faster than symbolic methods which explicitly compute invariants or covariants of the relevant action. We stress that (like almost all past algorithms) our convergence rate is polynomial in the approximation parameter; it is an intriguing question to achieve exponential convergence rate, beating symbolic algorithms exponentially, and providing *strong* membership and separation oracles for the problems above.

We strengthen and generalize the alternating minimization approach of previous papers by introducing the theory of highest weight vectors from representation theory into the numerical optimization framework. We show that highest weight vectors are natural potential functions for scaling algorithms and prove new bounds on their evaluations to obtain polynomial-time convergence. Our techniques are general and we believe that they will be instrumental to obtain efficient algorithms for moment polytopes beyond the ones consider here, and more broadly, for other optimization problems possessing natural symmetries. This is an extended abstract. Please see the full version of the paper [?] for technical details.

***Keywords*-Tensor scaling; quantum marginal problem; moment polytopes; invariant theory**

*PB is partially supported by DFG grant BU 1371 2-2. CF is supported in part by Simons Foundation award 332622. MW acknowledges support by the NWO through Veni grant no. 680-47-459. AW is partially supported by NSF grant CCF-1412958.

I. Introduction

A. Moment polytopes

As this paper is quite technical and contains some non-standard material for computer scientists, we begin with motivating the main object we study, as it is extremely natural from an optimization perspective: the *moment polytope*. Consider first the following diverse set of problems, trying to pick up common features among them (besides the obvious guess that they all are special cases of the framework we consider in this paper).

1) **The Schur-Horn Theorem:** Can a given Hermitian matrix be conjugated by unitary matrices to achieve a given diagonal?
2) **Eigenvalues of sums of Hermitian matrices:** Do there exist Hermitian $n \times n$ matrices A, B, C with prescribed eigenvalues such that $A + B = C$?
3) **Optimization:** Can a given non-negative matrix be converted to another with prescribed row and column sums, by only reweighing its rows and columns?

4) **Quantum information:** Can multiple parties, each holding a particle of a pure quantum state, locally transform their particles so that each particle is maximally entangled with the others?
5) **Analytic inequalities:** Given m linear maps $A_i : \mathbb{R}^n \to \mathbb{R}^{n_i}$ and $p_1, \ldots, p_m \geqslant 0$, does there exist a finite constant C such that for all integrable functions $f_i : \mathbb{R}^{n_i} \to \mathbb{R}_+$ we have

$$\int_{x \in \mathbb{R}^n} \prod_{i=1}^{m} f_i(A_i x) dx \leqslant C \prod_{i=1}^{m} \|f_i\|_{1/p_i}?$$

An important special case of such framework[1] is Cauchy-Schwarz, with $p_1 = p_2 = 1/2, m = 2, n = n_1 = n_2 = 1, C = 1, A_i = 1$.
6) **Algebraic complexity:** Given an arithmetic formula (with inversion gates) in non-commuting variables, is it non-zero?
7) **Polynomial support:** Given oracle access to a homogeneous polynomial p with non-negative integer coefficients on n variables, is a specified monomial (given as integer vector of exponents) in the Newton polytope[2] of p?

Some of the problems above are in P and for others, there are sufficient hints that they are in P (see [1–6]). While they may seem non-linear in their inputs, convexity plays an important role in each of them, as they all reduce to solving linear programs (implicitly defined with large number of facets). More specifically, each input to each problem defines a point and a polytope, and the answer is yes iff the point is in the polytope. These polytopes turn out to be special cases of *moment polytopes*.

This appearance of linearity and convexity is quite surprising, in some settings more so than others. Indeed, moment polytopes arise (and are used to understand problems) in many diverse settings such as symplectic geometry, algebraic geometry, lattice theory and others [7, 8]. The snag is that these polytopes are often defined by a huge number of inequalities (e.g. see [1]); typically the number is exponential or larger in the dimension of the input.[3] This motivates our efforts to develop efficient algorithms for them.

In order to explain the appearance of convex polytopes in these settings, we need to notice another common aspect of all problems above: their answers remain invariant under some group action! This is easy to see in some of the examples, which explicitly specify the groups. In the first, for matrices of size n, it is $U(n)$, the group of transformations conjugating the input. In the second, each of the three matrices may be conjugated by a unitary. In the 3rd, it is the product $\mathrm{T}(n) \times \mathrm{T}(n)$ of two (positive) diagonal invertible matrices which scale (resp.) the rows and columns. In the 4th problem, as each party is allowed to perform quantum measurements with post selection, the group representing each party's operations is $\mathrm{GL}(n)$ if its particle has n states, and so the full group is a direct product of these $\mathrm{GL}(n)$'s. The 5th problem is invariant to basis changes in the host space $\mathbb{R}^n$ and the other m spaces. The 6th is much harder to guess without Cohn's characterization of the free skew field, but turns out to be $\mathrm{GL}(n) \times \mathrm{GL}(n)$ acting on a different representation of the formulas. In the 7th, though it may not seem useful at first sight, $\mathrm{T}(n)$ acts by simply scaling every variable of the polynomial by a nonzero constant factor.

Having mentioned the two common features of the problems above (convexity and the invariance under a group action) we will now illustrate how one can use the structure of the group action in order to obtain moment polytopes. Let G be a "nice" [4] group acting linearly and continuously on a complex inner product space V and let v be a point in V. The *orbit* of a point $v \in V$ is the set of all vectors obtained by the action of G on v. The *orbit closure* of v is simply the closure of its orbit in the topology induced by the norm. As the previous paragraph observed, all of the problems above are questions about the orbit closures, which suggests understanding orbit closures is a fundamental task with many applications. A natural approach to study such orbit closures is by looking at the infinitesimal action of the group on every point

[1] These inequalities are the celebrated Brascamp-Lieb inequalities, which capture many more important inequalities such as Hölder's, Loomis-Whitney, and many others. See for instance [1] for a more detailed discussion.

[2] Given a polynomial $p(x_1, \ldots, x_n)$, define its support as the set of monomials whose coefficient in p is nonzero. The Newton polytope of $p(x_1, \ldots, x_n)$ is given by the convex hull of the exponent vectors of these monomials.

[3] However, in many of these areas even finiteness provides progress, as even decidability may not be obvious.

[4] The technical definition requires the group to be reductive (i.e. G is algebraic and every representation of G decomposes into a direct sum of irreducibles), connected, and that the inner product on V is an inner product that is invariant under the maximal compact subgroup of G. For the purpose of this paper, one can think of groups like $\mathrm{GL}(n)$, $\mathrm{T}(n)$, their direct products etc.

v.

This brings us to the *moment map*, denoted by $\mu_G(v)$, which is essentially a gradient of the log of the norm of v along the group action.[5] More explicitly, for each point v we can define the function $f_v(g) = \log\|g \cdot v\|_2^2$, and $\mu_G(v)$ will be the gradient of $f_v(g)$ evaluated at the identity element of G.[6] The moment map carries useful information about the group action, and one of its striking features is that the set of possible spectra of the image of any orbit closure under the moment map is a rational convex polytope [9–12]! That is a mouthful, so let us consider an example. Consider the action of $G = \mathrm{GL}(n)$ on some vector space V. Then the moment map maps V to $\mathrm{Mat}_{n\times n}(\mathbb{C})$ (set of all $n \times n$ complex matrices). Then the collection $\mathrm{spec}(\mu_G(v))$, as v varies over an orbit-closure forms a rational convex polytope. Here $\mathrm{spec}(M)$ denotes the vector of eigenvalues of M arranged in decreasing order. Note that $\mu_G(v)$ is a quadratic function of v, so the appearance of convexity is extremely surprising and non trivial. This polytope, which we will more explicitly see in the next section, is the so called *moment polytope* of the group action G on the orbit of v.

In the matrix scaling case (Problem 3), it turns out that the moment map applied to a certain matrix A gives us precisely the marginals of A (that is, the vector of row sums and column sums normalized to sum 1).[7] Thus, testing whether A can be scaled to another matrix with prescribed row and column sums is equivalent to testing whether the prescribed vector of row and column sums belongs to the moment polytope of the action of $\mathrm{T}(n)\times\mathrm{T}(n)$ on A. Similarly, all of the seven problems listed above fit into this framework (membership in moment polytope) for a suitable choice of group and representation.[8]

The reader might notice the dual nature of the problems above. They are both of algebraic as well as analytic nature. This phenomenon is extremely general and crucial for our paper. The analytic nature helps in designing algorithms, making the problem amenable to general optimization techniques, while the algebraic helps with analysis of these analytic algorithms and provides potential functions to track progress made by these algorithms. We will see that this will be the case for us as well.

B. Our setting

In this paper, we will be concerned with the action of the group of "local basis changes" on tensors, which are of interest for several reasons. The moment polytopes in this setting capture fundamental problems in quantum many-body physics - the so called *one-body quantum marginal problem*. They also capture fundamental problems in representation theory related to Kronecker coefficients, which are central objects of study in algebraic combinatorics and play an important role in geometric complexity theory. Moreover, as we will see, these moment polytopes generalize many of the settings described above and we believe that their complexity is representative of the complexity of general moment polytopes.

These moment polytopes (and their related problems) are most natural to state from the point of view of *quantum systems* and their *quantum marginals*[9], so we begin with these definitions. But before we define quantum systems some brief notation must be established.

Let $\mathrm{Ten}(n_0; n_1, \ldots, n_d) = \mathbb{C}^{n_0} \otimes \mathbb{C}^{n_1} \otimes \ldots \otimes \mathbb{C}^{n_d}$ denote the space of $d+1$ dimensional tensors of format $n_0 \times n_1 \times \cdots \times n_d$, and let X be a tensor in $\mathrm{Ten}(n_0; n_1, \ldots, n_d)$. If we regard X as a vector in $\mathbb{C}^{n_0 n_1 \ldots n_d}$, with $X^\dagger$ being it's conjugate transpose, then $\rho_X = XX^\dagger$ is a Hermitian positive semidefinite (PSD) operator on $\mathrm{Ten}(n_0; n_1, \ldots, n_d)$. We will denote by $\|X\| = \mathrm{tr}[\rho_X]^{1/2}$ the ℓ_2 norm of X (when viewed as a vector). With this notation in mind, we then define a *quantum system* with $d+1$ subsystems as a PSD operator on $\mathrm{Ten}(n_0; n_1, \ldots, n_d)$ with unit

[5] Indeed, the original name was *momentum map*, and is inspired from Hamiltonian physics, in which momentum is proportional to the derivative of position. Apparently moment maps are common in physics, where they are used to obtain conserved quantities (i.e. invariants) from symmetries of the phase space of symplectic manifolds describing some Hamiltonian system. In the general setting, we have the action of continuous group on a manifold, and the moment map provides a reverse map, from the manifold to the group (or more precisely, to the dual of the Lie algebra of the group).

[6] In this paper we assume G is a linear algebraic group, so we can take the gradient at the identity to mean the gradient of $f_v(e^A)$ at $A = 0$ in the vector space of matrices tangent to G at the identity.

[7] There is a slight technicality here and the moment map is actually the absolute values squared of the entries of A.

[8] For some of the problems mentioned above, it is non-trivial to phrase them as moment polytopes.

[9] These generalize the classical notion of marginals of a probability distribution on several variables.

trace[10].

Given a quantum system ρ on $\mathrm{Ten}(n_0; n_1, \ldots, n_d)$ and a subset $I \subseteq \{0, 1, \ldots, d\}$, we define its *(quantum) marginals* or *reduced density matrices* by $\rho^{(I)} = \mathrm{tr}_{I^c}[\rho]$, where tr_{I^c} denotes the partial trace over tensor factors $I^c = \{0, \ldots, d\} \backslash I$. In the same way that ρ describes the state of the entire quantum system, $\rho^{(I)}$ characterizes the state of the subsystems labeled by I (in an analogous way to the classical marginal of a probability distribution). For $I = \{i\}$, we write $\rho^{(i)}$; these operators are known as the *one-body marginals* or *one-body reduced density matrices* of ρ. Each $\rho^{(i)}$ is uniquely characterized by the property that

$$\mathrm{tr}[\rho^{(i)} A^{(i)}] = \mathrm{tr}[\rho(I_{n_0} \otimes I_{n_1} \otimes \ldots \otimes I_{n_{i-1}} \otimes A^{(i)} \otimes \ldots \otimes I_{n_d})] \tag{1}$$

for all operators $A^{(i)}$ on $\mathbb{C}^{n_i}$.

For a tensor $X \in \mathrm{Ten}(n_0; n_1, \ldots, n_d)$ and a given subset $I \subseteq \{0, 1, \ldots, d\}$ of the subsystems, the marginals of ρ_X with respect to I have a particularly simple description: using the standard basis, identify X with a matrix $M_X^{(I)} \in \mathrm{Mat}_{n_I, n_{I^c}}(\mathbb{C})$, where we denote $n_I := \prod_{i \in I} n_i$. The matrix $M_X^{(I)}$ is known as a flattening, unfolding, or matricization [13, 14] of the tensor X. Then, $\rho_X^{(I)} = M_X^{(I)} (M_X^{(I)})^\dagger$ is its Gram matrix.

Given a Hermitian operator σ on $\mathbb{C}^n$ (i.e., an $n \times n$ Hermitian matrix), we write $\mathrm{spec}(\sigma) = (s_1, \ldots, s_n)$ for the vector of eigenvalues of σ, ordered non-increasingly. If σ is PSD with unit trace then its eigenvalues form a probability distribution, so $\mathrm{spec}(\sigma)$ is an element of $P_+(n) := \{(s_1, \ldots, s_n) : s_1 \geqslant \cdots \geqslant s_n \geqslant 0 : \sum_j s_j = 1\}$. We also abbreviate $P_+(n_1, \ldots, n_d) := P_+(n_1) \times \cdots \times P_+(n_d)$. We will be particularly interested in characterizing the eigenvalues of the one-body marginals, motivated by the following fundamental problem in quantum mechanics [15]:

Problem I.1 (One-body quantum marginal problem)**.** *Given $\boldsymbol{p} \in P_+(n_1, \ldots, n_d)$, decide if there exists a tensor $Y \in \mathrm{Ten}(1; n_1, \ldots, n_d)$ such that $\mathrm{spec}(\rho_Y^{(i)}) = \boldsymbol{p}^{(i)}$ for all $i = 1, \ldots, d$.*

Remark I.2. *Note that the above problem is equivalent to the following, given density matrices (PSD matrices with unit trace) $\rho^{(1)}, \ldots, \rho^{(d)}$, determine if there exists a tensor (pure state) $Y \in \mathrm{Ten}(1; n_1, \ldots, n_d)$ such that $\rho_Y^{(i)} = \rho_i$ for all $i = 1, \ldots, d$. Since a unitary change of basis comes for free on each subsystem, only the eigenvalues of $\rho^{(1)}, \ldots, \rho^{(d)}$ are relevant.*

The above problem is extremely fundamental from the point of view of quantum many-body physics. It is a special case of the more general quantum marginal problem, which puts constraints on the marginals of multiple systems and is known to be QMA-complete (for growing d) [16].

We note that the normalization to trace one is natural; since $\mathrm{tr}[\rho_Y^{(i)}] = \|Y\|^2$ for all i, we can simultaneously rescale all marginals simply by rescaling the tensor.

Now we discuss how Problem I.1 can be phrased as a question about moment polytopes [15, 17–21]. Let $G = \mathrm{GL}(n_1) \times \cdots \times \mathrm{GL}(n_d)$, where $\mathrm{GL}(n)$ denotes the group of invertible $n \times n$-matrices. Then G acts on $V = \mathrm{Ten}(n_0; n_1, \ldots, n_d)$ by

$$(g^{(1)}, \ldots, g^{(d)}) \cdot X := (I_{n_0} \otimes g^{(1)} \otimes \ldots \otimes g^{(d)}) X.$$

As the group acts by rescaling slices of the tensor, we will call any $Y \in G \cdot X$ a *tensor scaling* of X.[11]

What is the moment map in this setting? It turns out that the moment captures exactly the notion of one-body quantum marginals. It is more convenient to define the moment map on the projective space, since we don't care about the scalar multiples. We will denote the projective space corresponding to V by $\mathbb{P}(V)$ and identify it with the set of rank-one trace-one PSD operators on V, $\mathbb{P}(V) = \{\rho = [X] = XX^\dagger / X^\dagger X : 0 \neq X \in V\}$. Then the moment map can be written as[12]

$$\mu \colon \mathbb{P}(V) \to \mathrm{Herm}(n_1) \times \cdots \times \mathrm{Herm}(n_d), \quad \rho \mapsto (\rho^{(1)}, \ldots, \rho^{(d)}), \tag{2}$$

where $\mathrm{Herm}(n)$ denotes the space of Hermitian $n \times n$-matrices. Now consider a projective subvariety $\mathcal{X}$ of $\mathbb{P}(V)$ such as $\mathcal{X} = \mathbb{P}(V)$ or an orbit-closure[13] i.e. $\mathcal{X} = \overline{G \cdot [X]}$ for some given tensor $X \in V$.[14] Let

[10] A reader not familiar with the basics of quantum systems may want to skip a couple of paragraphs ahead.

[11] The extra coordinate with dimension n_0 can be equivalently thought of as enumerating an n_0-tuple of tensors in $\mathrm{Ten}(n_1, \ldots, n_d)$ and the group G acts simultaneously on all the tensors in the tuple. Much of the theory remains similar if one sets $n_0 = 1$ and that can be done mentally on a first reading. In the quantum language, it is the difference between acting on pure states ($n_0 = 1$) vs acting on mixed states ($n_0 > 1$).

[12] After identifying the Lie algebra of K with its dual.

[13] Here, the closure can be taken either in the Euclidean or in the Zariski topology.

[14] In general $\mathcal{X}$ can be any G-stable irreducible projective subvariety of $\mathbb{P}(V)$.

us look at the collection of marginal eigenvalues when restricted to tensors in $\mathcal{X}$:

$$\Delta(\mathcal{X}) := \{(\mathrm{spec}(\rho^{(1)}), \ldots, \mathrm{spec}(\rho^{(d)})) : \rho \in \mathcal{X}\} \subseteq P_+(n_1, \ldots, n_d). \quad (3)$$

We emphasize again the amazing, surprising and non-trivial fact that $\Delta(\mathcal{X})$ is a rational convex polytope [12, 17, 18, 22] – known as the *moment polytope* or *Kirwan polytope* of $\mathcal{X}$.[15] This means that $\Delta(\mathcal{X})$ can in principle be given in terms of finitely many affine inequalities in eigenvalues of the one-body marginals [15, 19, 20]. In particular, the preceding applies to $\mathcal{X} = \mathbb{P}(V)$, so we can rephrase Problem I.1 as follows: *Given* $\boldsymbol{p} \in P_+(n_1, \ldots, n_d)$, *is it a point in* $\Delta(1; n_1, \ldots, n_d) := \Delta(\mathbb{P}(V))$*?* More generally, we can consider the following decision problem:

Problem I.3 (General moment polytope)**.** *Given* $\boldsymbol{p} \in P_+(n_1, \ldots, n_d)$*, decide if there exists a tensor* $[Y] \in \mathcal{X}$ *such that* $\mathrm{spec}(\rho_Y^{(i)}) = \boldsymbol{p}^{(i)}$ *for all* $i = 1, \ldots, d$*.*

When $\mathcal{X} = \overline{G \cdot [X]}$ is the *orbit closure* of some given tensor $X \in V$, we will abbreviate the moment polytope by $\Delta(X) := \Delta(\overline{G \cdot [X]})$. In quantum information theory, moment polytopes of orbit closures have been called *entanglement polytopes* as they characterize the multipartite entanglement from the perspective of the one-body marginals [23, 24]. But, along with the corresponding invariant-theoretic multiplicities, they are also of interest in algebraic and geometric complexity theory [25–28]. The corresponding decision problem is the following:

Problem I.4 (Moment polytope of orbit closure)**.** *Given* $X \in \mathrm{Ten}(n_0; n_1, \ldots, n_d)$ *and* $\boldsymbol{p} \in P_+(n_1, \ldots, n_d)$*, decide if there exists* $Y \in \overline{G \cdot X}$ *such that* $\mathrm{spec}(\rho_Y^{(i)}) = \boldsymbol{p}^{(i)}$ *for all* $i = 1, \ldots, d$*.*

That is, Problem I.4 asks whether $\boldsymbol{p} = (\boldsymbol{p}^{(1)}, \ldots, \boldsymbol{p}^{(d)})$ is a point in $\Delta(X)$.[16]

One can show that Problem I.4 is intimately related to Problems I.1 and I.3: $\boldsymbol{p} \in \Delta(\mathcal{X})$ iff $\boldsymbol{p} \in \Delta(X)$ for a generic $X \in \mathcal{X}$ (see the full version). This is explained in the full version of the paper. We will therefore focus our attention on Problem I.4.

It is natural to go beyond the decision problem and look for an algorithm that finds a tensor Y with the desired marginals, as well as the group element that transforms X into Y. Since such an Y will be in the orbit through X, we demand only that the marginals are correct up to some target accuracy.

Definition I.5 (ε-close)**.** *The marginals of* $Y \in \mathrm{Ten}(n_0; n_1, \ldots, n_d)$ *are said to be* ε-close *to* $\boldsymbol{p} \in P_+(n_1, \ldots, n_d)$ *if* $\|\mathrm{spec}(\rho_Y^{(i)}) - \boldsymbol{p}^{(i)}\|_1 \leqslant \varepsilon$ *for* $i = 1, \ldots, d$*. Here,* $\|\boldsymbol{x}\|_1 = \sum_j |x_j|$ *is the* ℓ^1*-norm.*

Problem I.6 (Tensor scaling)**.** *Given* $X \in \mathrm{Ten}(n_0; n_1, \ldots, n_d)$*,* $\boldsymbol{p} \in \Delta(X)$*, and* $\varepsilon > 0$*, find* $g_\varepsilon \in G$ *such that* $Y = g_\varepsilon \cdot X$ *has marginals that are* ε*-close to* $\boldsymbol{p}$*.*

While it may not be immediately clear, there exist scalings as in Problem I.6 for any $\varepsilon > 0$ if and only if the answer to Problem I.4 is yes, i.e., if and only if $\boldsymbol{p} \in \Delta(X)$.

The polytopes $\Delta(\mathcal{X})$ admit alternative characterization in terms of invariant theory [17]. This connection is explained in the full version of the paper, and it is central to the analysis of our algorithms. For now, we only mention an important special case. Let $g(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\nu})$ denote the *Kronecker coefficients*, which are fundamental objects in the classical representation theory of the symmetric and general linear groups [30, 31]. They also feature in geometric complexity theory as a potential way of creating representation theoretic obstructions [25, 32]. For example, $g(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\nu})$ can be defined as the multiplicity of the irreducible S_k-representation $[\boldsymbol{\lambda}]$ in the tensor product $[\boldsymbol{\mu}] \otimes [\boldsymbol{\nu}]$. Here, $\boldsymbol{\lambda}$, $\boldsymbol{\mu}$, and $\boldsymbol{\nu}$ are

[15] Note that we have identified $\mathbb{P}(V)$ with the set of rank 1 density matrices and hence it is far from being a convex set - yet the spectrum of its image under the moment map is convex.

[16] When $n_0 = 1$, there is a physical interpretation of the orbit-closure. $Y \in \overline{G \cdot X}$ means that Y can be obtained to arbitrary precision from X (which is naturally understood as a d-partite quantum state) by a class of quantum operations known as *stochastic local operations and classical communication (SLOCC)* [29]. SLOCC can be intuitively understood as follows: we imagine that different parties hold the different systems of a quantum state; SLOCC then corresponds to a sequence of local quantum operations and measurements, where we allow for *post-selection* on specific measurement outcomes. Problem I.4 then asks if given a tensor $X \in \mathrm{Ten}(n_1, \ldots, n_d)$, does there exist a Y obtainable by a sequence of SLOCC operations from X s.t. $\mathrm{spec}(\rho_Y^{(i)}) = \boldsymbol{p}^{(i)}$ for all i. This is a generalization of the SLOCC entanglement distillation question where $\boldsymbol{p}^{(i)}$ is the uniform distribution for all i.

partitions, which we may think of nonincreasing vectors in $\mathbb{Z}^n_{\geqslant 0}$ with $\sum_j \lambda_j = \sum_j \mu_j = \sum_j \nu_j =: k$. Then,

$$\exists \text{ integer } s \geqslant 1 : g(s\boldsymbol{\lambda}, s\boldsymbol{\mu}, s\boldsymbol{\nu}) > 0 \quad \Leftrightarrow \quad \frac{1}{k}(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\nu}) \in \Delta(1; n, n, n), \tag{4}$$

so the solution to the one-body quantum marginal problem captures precisely the *asymptotic support* of the Kronecker coefficients [15, 33, 34]. We note that the problem of deciding whether $g(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\nu}) > 0$ is known to be NP-hard [35]. However since the asymptotic vanishing of Kronecker coefficients is captured by the quantum marginal problem, it has been conjectured that it should have a polynomial time algorithm and we make progress towards this question.[17] Since Kronecker coefficients are so poorly understood, understanding their asymptotic support would also go a long way in understanding them.

C. Prior work

As mentioned above, Problem I.1 can be approached by first computing (the defining inequalities of) the moment polytope $\Delta(n_0; n_1, \ldots, n_d)$. The problem of computing moment polytopes has a long history in mathematics (e.g., [12, 19, 22, 38–43]). That the one-body quantum marginal problem falls into this framework was first noticed by Klyachko [44], who gave a complete description of the polytopes in terms of finite lists of inequalities (cf. [15, 45, 46]). Before that, only low-dimensional special cases were known [47–49]. Further developments include the minimal complete description from [19] and the cohomology-free variant [20]. Yet, all these descriptions in terms of inequalities are largely computationally infeasible; explicit descriptions are known up to formats $3 \times 3 \times 9$ [15] and $4 \times 4 \times 4$ [20], and when all dimensions are two [47].

Problems I.3 and I.4 can in principle be approached using classical computational invariant theory (e.g., [23, 50, 51]), based on the invariant-theoretic description of $\Delta(\mathcal{X})$ and degree bounds (see full version). In practice, however, this is completely infeasible except for very small dimensions. The problem of describing $\Delta(\mathcal{X})$ also falls into the framework of [19], but it is not clear how to turn this into an algorithm. In summary, all the methods described above are computationally expensive and take time at least exponential in the input size.

None of the preceding algebraic methods can be used to solve Problem I.6, since they only decide membership but do *not* produce the transformation that produces a tensor with the desired target spectra. This calls for the development of numerical algorithms for Problem I.6. Curiously, this development stemmed from motivations in algebraic complexity and the PIT problem. The first such algorithm was proposed in [52]. Its complexity analysis, that brought on the connection to invariant theory (and other fields, some mentioned above) was achieved in [5]. In the language we use here, it deals with $d = 2$ (operator scaling) and uniform marginals, and results in polynomial time algorithms for problems in diverse areas discussed there.[18] The operator scaling problem was then extended in two directions, which we mention next: one direction being general values of d (tensor scaling) and the other being $d = 2$ and arbitrary marginals.

For general d, a deterministic algorithm was given in [4] (based on a proposal in [55] for $n_0 = 1$). Very recently, a randomized polynomial time algorithm for operator scaling to general marginals was given in [56]. The two papers [4, 56] study two different generalizations of the operator scaling problem in [5]. The present paper completes a natural square by studying a common generalization of the problems studied in [4, 56]. All these algorithms can be seen as noncommutative generalizations of the Sinkhorn-Knopp algorithm for 'matrix scaling' [57, 58].

It was recently shown that Problem I.1 is in NP∩coNP [59]. In view of Eq. (4), this should be contrasted with the NP-hardness of deciding whether a single Kronecker coefficient is zero or not [35].

D. Summary of results

Our main result in this paper is a randomized algorithm for tensor scaling to general marginals (Problem I.6). As a consequence, we obtain algorithms for all other problems.

Theorem I.7. *There is a randomized algorithm running in time* $\mathrm{poly}(N, 1/\varepsilon)$*, that takes as input* $X \in$

[17] We note that the closely related Littlewood-Richardson coefficients (which capture the same problem for the representations of the general linear group) satisfy the so called saturation property: $c(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\nu}) > 0$ iff $c(s\boldsymbol{\lambda}, s\boldsymbol{\mu}, s\boldsymbol{\nu}) > 0$ [36]. Hence the asymptotic support is the same as support for this case and this is also a key ingredient in the polynomial time algorithms for testing if $c(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\nu}) > 0$ [37].

[18] The underlying algebraic problem associated with operator scaling, namely non-commutative singularity and rank of symbolic matrices found a different, algebraic algorithm in the works of [53, 54]

$\mathrm{Ten}(n_0; n_1, \ldots, n_d)$ *with Gaussian integer entries (specified as a list of real and complex parts, each encoded in binary, with bit size* $\leqslant b$*) and* $\boldsymbol{p} \in P_+(n_1, \ldots, n_d)$ *with rational entries (specified as a list of numerators and denominators, each encoded in binary, with bit size* $\leqslant b$*). The algorithm either correctly identifies that* $\boldsymbol{p} \notin \Delta(X)$*, or it outputs a scaling* $g \in G$ *such that the marginals of* $g \cdot X$ *are* ε*-close to the target spectra* $\boldsymbol{p}$*. Here* N *is the total bit-size of the input,* $N = 2n_0 n_1 \cdots n_d b + 2(n_1 + \cdots n_d)b$*.*

As a consequence of Theorem I.7, we obtain a randomized algorithm for a promise version of the membership Problem I.4 (and hence for Problem I.1, see full version).

Corollary I.8. *There is a randomized algorithm running in time* $\mathrm{poly}(N, 1/\varepsilon)$*, that takes as input* $X \in \mathrm{Ten}(n_0; n_1, \ldots, n_d)$ *with Gaussian integer entries (specified as a list of real and complex parts, each encoded in binary, with bit size* $\leqslant b$*) and* $\boldsymbol{p} \in P_+(n_1, \ldots, n_d)$ *with rational entries (specified as a list of numerators and denominators, each encoded in binary, with bit size* $\leqslant b$*). The algorithm distinguishes between the following two cases:*

1) $\boldsymbol{p} \in \Delta(X)$*.*
2) $\boldsymbol{p}$ *is* ε*-far (in* ℓ_1*-norm) from any point* $\boldsymbol{q} \in \Delta(X)$*.*

Here N *is the total bit-size of the input,* $N = 2n_0 n_1 \cdots n_d b + 2(n_1 + \cdots n_d)b$*.*

This yields the following corollary.

Corollary I.9. *There is a randomized algorithm running in time* $\mathrm{poly}(n_0 n_1 \cdots n_d, b, 1/\varepsilon)$*, that takes as input* $\boldsymbol{p} \in P_+(n_1, \ldots, n_d)$ *with rational entries (specified as a list of numerators and denominators, each encoded in binary, with bit size* $\leqslant b$*). The algorithm distinguishes between the following two cases:*

1) $\boldsymbol{p} \in \Delta(1; n_1, \ldots, n_d)$*, i.e., there exists* $Y \in \mathrm{Ten}(n_1, \ldots, n_d)$ *such that* $\mathrm{spec}(\rho_Y^{(i)}) = \boldsymbol{p}^{(i)}$ *for all* i*.*
2) $\boldsymbol{p}$ *is* ε*-far (in* ℓ_1*-norm) from any point* $\boldsymbol{q} \in \Delta(1; n_1, \ldots, n_d)$*.*

As described before, Problem I.1 captures the asymptotic vanishing of Kronecker coefficients. Hence we get the following corollary which describes a randomized polynomial time algorithm for a promise version of the asymptotic Kronecker problem.

Corollary I.10. *There is a randomized algorithm running in time* $\mathrm{poly}(n, b, 1/\varepsilon)$*, that takes as input three partitions* $\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\nu} \in \mathbb{Z}^n_{\geqslant 0}$ *with entries described in binary with bit-size at most* b*. The algorithm distinguishes between the following two cases:*

1) *There exists an integer* $s \geqslant 1$ *s.t.* $g(s\boldsymbol{\lambda}, s\boldsymbol{\mu}, s\boldsymbol{\nu}) > 0$*.*
2) *For all* $\boldsymbol{\lambda}', \boldsymbol{\mu}', \boldsymbol{\nu}'$ *s.t.* $g(\boldsymbol{\lambda}', \boldsymbol{\mu}', \boldsymbol{\nu}') > 0$*, it holds that* $(\boldsymbol{\lambda}', \boldsymbol{\mu}', \boldsymbol{\nu}') / |(\boldsymbol{\lambda}', \boldsymbol{\mu}', \boldsymbol{\nu}')|$ *is* ε*-far (in* ℓ_1*-norm) from* $(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\nu})/|(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\nu})|$*.*

Here g *denotes the Kronecker coefficient and* $|(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\nu})| = \sum_j \lambda_j = \sum_j \mu_j = \sum_j \nu_j$*.*

In many applications, the tensor X can be more succinctly represented than by its $n_0 n_1 \cdots n_d$ many coordinates. If the representation is preserved by scalings and allows for efficient computation of the marginals, then this yields a useful improvement to Algorithm 1. A prime example of which are the so called *matrix-product states* or *tensor-train decompositions* with polynomial bond dimension [55, 60]. We won't define these states here (see full version for a formal definition) but we will just say that these have much smaller (exponentially smaller in d) descriptions than specifying all the $n_0 n_1 \cdots n_d$ coordinates of the tensors. This class includes the *unit tensors* and the *matrix multiplication tensors*, which are central objects in algebraic complexity theory [26, 61] and whose moment polytopes are *not* known!

Theorem I.11 (Informal)**.** *There is a randomized algorithm running in time* $\mathrm{poly}(N, b, 1/\varepsilon)$*, that takes as input a matrix-product state* $X \in \mathrm{Ten}(n_0; n_1, \ldots, n_d)$ *with input size* N *and* $\boldsymbol{p} \in P_+(n_1, \ldots, n_d)$ *with rational entries (specified as a list of numerators and denominators, each encoded in binary, with bit size* $\leqslant b$*). The algorithm either correctly identifies that* $\boldsymbol{p} \notin \Delta(X)$*, or it outputs a scaling* $g \in G$ *such that the marginals of* $g \cdot X$ *are* ε*-close to the target spectra* $\boldsymbol{p}$*.*

It is a very exciting open problem to improve the running time dependence on ε in Corollary I.8, Corollary I.9 and Corollary I.10 to $\mathrm{poly}(\log 1/\varepsilon)$. This would yield randomized polynomial time algorithms for Problem I.1, Problem I.4 and the asymptotic Kronecker problem due to the following theorem that we prove in the full version.

Theorem I.12 (Minimal gap)**.** *Let* X *be a nonzero tensor in* $\mathrm{Ten}(n_0; n_1, \ldots, n_d)$*. If* $[Y] \in \overline{G \cdot [X]}$ *is a scaling with marginals that are* $\gamma(n_1, \ldots, n_d, \ell)$*-close to* $\boldsymbol{p}$*, then* $\boldsymbol{p} \in \Delta(\mathcal{X})$*. Here* $\gamma(n_1, \ldots, n_d, \ell) = \exp\left(-O\left((n_1 + \cdots + n_d)\log(\ell \max_j n_j)\right)\right)$ *and* ℓ *is the minimal integer s.t.* $\ell \boldsymbol{p}$ *has integral entries.*

An analogous result for the full moment polytope

$\Delta(\mathbb{P}(V))$ was proven in [59]. We believe that the inverse exponential bound in the above theorem cannot be improved to an inverse polynomial bound. Therefore developing scaling algorithms with runtime dependence $\mathrm{poly}(\log 1/\varepsilon)$ is of paramount importance.

Before describing our algorithm and the high level intuition for its analysis, let us describe the algorithm and analysis for a rather special case of matrix scaling, which turns out to very enlightening.

E. Simple example: matrix scaling

The matrix scaling problem (Problem 3 in Section I-A) provides us with a template for what is to come, and understanding the evolution of a particular algorithm for this problem will give us intuition on how to solve the more general tensor scaling problem, and how invariant theory naturally appears.

If one wants to scale a given $n \times n$ matrix A to a doubly stochastic matrix (that is, one whose rows and columns each sum to 1), a natural algorithm (first proposed in [57]) arises from the fact that the group is a Cartesian product. We can alternately use scalings of the form $(R, I) \in \mathrm{T}(n) \times \mathrm{T}(n)$ to normalize the row sums of A and scalings of the form $(I, C) \in \mathrm{T}(n) \times \mathrm{T}(n)$ to normalize the column sums of A.

To this end, set $R(A)$ to be a diagonal matrix having $R(A)_{i,i}$ to be the inverse of the sum of the elements of the i^{th} row of A, and define $C(A)$ in a similar way for the columns of A. The algorithm can be described as follows: repeatedly (for a polynomial number of iterations) apply the following steps:

- Normalize the rows of A. That is, $A \leftarrow R(A) \cdot A$
- Normalize the columns of A. That is, $A \leftarrow A \cdot C(A)$.

If, throughout this process, the matrix A never gets sufficiently close to a doubly stochastic matrix (in ℓ_2-distance), then we will conclude that A cannot be scaled to doubly stochastic; otherwise we can conclude that A can be scaled to doubly stochastic. The process also gives us a way to *obtain* the scalings that approach doubly stochastic - while there are multiple algorithms for the decision problem (which turns out to be the bipartite perfect matching problem), not all help find the scalings!

The analysis of this algorithm (from [3]; also see [62] for a different potential function) is extremely simple, and follows a three step approach based on a progress measure $P(A) = \mathrm{Perm}(A)$.
The following two properties of the potential function will be useful for us.

1) If A is scalable to doubly stochastic, then $P(A) > 0$.
2) $P(A) \leqslant 1$ if A is row- or column-normalized.

The three step approach then is the following:

1) **[Lower bound]**: Initially $P(A) > 2^{-\mathrm{poly}(n)}$ (w.l.o.g. we assume A is row-normalized) [19].
2) **[Progress per step]**: If A is row- or column-normalized and sufficiently far from being doubly stochastic, then normalizing A increases $P(A)$. One can explicitly bound the increase using a robust version of the AM-GM inequality.
3) **[Upper bound]**: $P(A)$ is bounded by 1 if A is row- or column-normalized.

This three-step analysis shows that the scaling algorithm is able to solve the doubly stochastic scaling problem in polynomial time. The difficult part of the analysis is coming up with a potential function satisfying the properties above. This is the role played by invariant theory later. A source of good potential functions will turn out to be *highest weight vectors*, which are (informally speaking) "eigenvectors" of the action of certain subgroups of the main group action. Note that the permanent is an eigenvector of the action of $\mathrm{T}(n) \times \mathrm{T}(n)$ since $\mathrm{Perm}(RXC)$ equals $\left(\prod_i R_{i,i} \cdot \prod_j C_{j,j}\right) \cdot \mathrm{Perm}(X)$ for $(R, C) \in \mathrm{T}(n) \times \mathrm{T}(n)$.

If we want to solve the more general scaling problem, where we are given a prescribed value for the row and column sums, say as an non-negative integer vector $(r, c) = (r_1, \ldots r_n, c_1, \ldots, c_n)$, the same natural algorithm can be applied. The only change one needs to make in the algorithm above is that we will now normalize the rows of A to have sums $(r_1, \ldots, r_n)$ and the columns to have sum $(c_1, \ldots, c_n)$. The analysis is also quite similar: one can choose the potential function, for example, to be the permanent of matrix B obtained from A by repeating i^{th} row r_i times and j^{th} column c_j times. However, the distinction between the uniform and the non-uniform versions of the problems is much starker in our higher dimensional non-commutative setting, as we will see next.

[19] There is some dependency on the bit complexity of the input that we are ignoring.

F. Techniques and proof overview

Our algorithm and its analysis generalize two recent works [4, 56], which in turn generalize the analysis of matrix scaling in Section I-E. The paper [4] studies the special case when $\boldsymbol{p}^{(i)}$ is the uniform distribution (over a set of size n_i) for all i while the paper [56] studies the special case $d = 2$. Our algorithm is a natural common generalization of the algorithms in [4, 56] while our analysis generalizes the analysis in [4] replacing the use of invariants with highest weight vectors (we will explain what these are later).

Let us develop some intuition for the algorithm. It is usually the case with scaling problems, as we saw with matrix scaling, and more generally in the framework of alternating minimization, that one of the constraints is easy to satisfy by scaling. The same is true for the problem we have at hand. We are given a tensor $X \in \mathrm{Ten}(n_0; n_1, \ldots, n_d)$. Suppose we want $\mathrm{spec}\left(\rho_X^{(i)}\right) = \boldsymbol{p}^{(i)}$. With the shorthand $\boldsymbol{p}_\uparrow^{(i)} := (p_{n_i}^{(i)}, \ldots, p_1^{(i)})$, we act on X by $g = \left(I, I, \ldots, \mathrm{diag}\left(\boldsymbol{p}_\uparrow^{(i)}\right)^{1/2} \left(\rho_X^{(i)}\right)^{-1/2}, \ldots, I\right)$, where the non-trivial element is in the i^{th} location. This will satisfy the i^{th} constraint. Or indeed, one can choose any matrix R s.t. $RR^\dagger = \rho_X^{(i)}$ and act on X by $g = \left(I, I, \ldots, \mathrm{diag}\left(\boldsymbol{p}_\uparrow^{(i)}\right)^{1/2} R^{-1}, \ldots, I\right)$. This will also satisfy the i^{th} constraint. By choosing each time to fix the index which is "farthest" from its target spectrum, we have defined an iterative algorithm (up to the choice of R at each step) that keeps on alternately fixing the constraints. It turns out that this algorithm works (for any choice of R at each step!) when the $\boldsymbol{p}^{(i)}$'s are all uniform and converges in a polynomial number of iterations [4].

Interestingly, the choice of R that works for general $\boldsymbol{p}^{(i)}$'s is that of upper triangular matrices![20] This was the choice made in [56] as well. This restriction on scaling factors will make the analysis more complicated as we shall soon see. One intuitive reason for the difference between the uniform and the general case is the following: in the general case, we made an arbitrary decision to try to scale X to have marginals $\mathrm{diag}\left(\boldsymbol{p}^{(i)}\right)$ while we could have chosen to scale it to any $\rho^{(i)}$ s.t. $\mathrm{spec}(\rho^{(i)}) = \mathrm{diag}\left(\boldsymbol{p}^{(i)}\right)$. This choice of basis is not present in the uniform case since all bases are the same!

This restriction on scaling factors creates another problem: it disconnects the orbit space (see example below). Thus, we need to initialize the algorithm with a random basis change of the given input, and only then resume the restricted scaling. This idea is used as well in [56]. We explain, via an example, why this random basis change (or at least a "clever" basis change) is needed at the start of the algorithm. Consider the diagonal unit tensor $X \in \mathrm{Ten}(1; 2, 2, 2)$, where $X_{j,k,\ell} = 1$ iff $j = k = \ell$. It is easy to see that without the initial randomization, the algorithm (which chooses an upper triangular R at each step) would only produce diagonal tensors Y ($Y_{j,k,\ell} \neq 0$ iff $j = k = \ell$). And the marginals of any such tensor are isospectral. On the other hand, the G-orbit of X is dense in $\mathrm{Ten}(1; 2, 2, 2)$ and so $\Delta(X) = \Delta(1; 2, 2, 2)$. In particular, X can be scaled to tensors with non-isospectral marginals.

The algorithm is described as Algorithm 1. The following is the main theorem regarding the analysis of Algorithm 1 from which Theorem I.7 follows up to an analysis of the bit complexity of Algorithm 1.

Theorem I.13 (Tensor scaling)**.** *Let X be a (nonzero) tensor in $\mathrm{Ten}(n_0; n_1, \ldots, n_d)$ whose entries are Gaussian integers of bitsize no more than b. Also, let $\boldsymbol{p} \in P_+(n_1, \ldots, n_d)$ with rational entries of bitsize no more than b such that $p_{n_i}^{(i)} > 0$ for all $i = 1, \ldots, d$. Finally, let $\varepsilon > 0$.*

Then, with probability at least 1/2, Algorithm 1 either correctly identifies that $\boldsymbol{p} \notin \Delta(X)$, or it outputs $g \in G$ such that the marginals of $Y = g \cdot X$ are ε-close to $\boldsymbol{p}$. In fact, we have

$$\|\rho_Y^{(i)} - \mathrm{diag}(\boldsymbol{p}_\uparrow^{(i)})\|_{\mathrm{tr}} \leqslant \varepsilon \quad \text{for } i = 1, \ldots, d \tag{5}$$

in the latter case, where $\|A\|_{\mathrm{tr}} = \mathrm{tr}[\sqrt{A^\dagger A}]$ is the trace norm.

Remark I.14. *Note that the condition*

$$\|\rho_Y^{(i)} - \mathrm{diag}\left(\boldsymbol{p}_\uparrow^{(i)}\right)\|_{\mathrm{tr}} \leqslant \varepsilon$$

[20] This choice works for all $\boldsymbol{p}^{(i)}$'s. We don't know if this choice of upper-triangularity is necessary. There is also a nice interpolation between the case of uniform $\boldsymbol{p}^{(i)}$'s and $\boldsymbol{p}^{(i)}$'s with distinct entries. See full version.

[21] Usually the Cholesky decomposition refers to $\rho = LL^\dagger$ where L is lower triangular. However using such a decomposition for a different matrix, one can easily obtain $\rho = RR^\dagger$, where R is upper triangular. Simply set $R = PLP$ where P is a permutation matrix which swaps i and $n - i$ and $P\rho P = LL^\dagger$, where L is lower triangular.

Input: $X \in \mathrm{Ten}(n_0; n_1, \ldots, n_d)$ with Gaussian integer entries (specified as a list of real and complex parts, each encoded in binary, with bit size $\leqslant b$) and $\boldsymbol{p} \in P_+(n_1, \ldots, n_d)$ with rational entries (specified as a list of numerators and denominators, each encoded in binary, with bit size $\leqslant b$) such that $p_{n_i}^{(i)} > 0$ for all $i = 1, \ldots, d$.

Output: Either the algorithm correctly identifies that $\boldsymbol{p} \notin \Delta(X)$, or it outputs $g \in G$ such that the marginals of $Y := g \cdot X$ satisfy Eq. (5); in particular the marginals are ε-close to the target spectra $\boldsymbol{p}$.

Algorithm:

1) Let $\ell > 0$ such that $\ell \boldsymbol{p}^{(i)}$ has integer entries for all $i = 1, \ldots, d$. Let $g = (g^{(1)}, \ldots, g^{(d)})$ denote the tuple of matrices ($g^{(i)}$ is $n_i \times n_i$) whose entries are chosen independently uniformly at random from $\{1, \ldots, M\}$, where $M := 2dK$ and $K := (\ell d \max_{i=1}^d n_i)^{d \max_{i=1}^d n_i^2}$.
2) For $i = 1, \ldots, d$, if the marginal $\rho_{g \cdot X}^{(i)}$ is singular then output $\boldsymbol{p} \notin \Delta(X)$ and halt. Otherwise, update $g^{(1)} \leftarrow g^{(1)} / \|g \cdot X\|$.
3) For $t = 1, \ldots, T$ where T is defined to be

$$\left\lceil \frac{32 \ln 2}{\varepsilon^2} \left(3 \sum_{i=0}^d \log_2(n_i) + b + d \log_2(M) \right) \right\rceil,$$

repeat the following:

- Compute $Y := g \cdot X$ and, for $i = 1, \ldots, d$, the one-body marginals $\rho_Y^{(i)}$ and the distances $\varepsilon^{(i)} := \|\rho_Y^{(i)} - \mathrm{diag}(\boldsymbol{p}_\uparrow^{(i)})\|_{\mathrm{tr}}$.
- Select an index $i \in \{1, \ldots, d\}$ for which $\varepsilon^{(i)}$ is largest. If $\varepsilon^{(i)} \leqslant \varepsilon$, output g and halt.
- Compute the Cholesky decomposition[21] $\rho_Y^{(i)} = R^{(i)} (R^{(i)})^\dagger$, where $R^{(i)}$ is an upper-triangular matrix. Update $g^{(i)} \leftarrow \mathrm{diag}(\boldsymbol{p}_\uparrow^{(i)})^{1/2} (R^{(i)})^{-1} g^{(i)}$.

4) Output $\boldsymbol{p} \notin \Delta(X)$.

Algorithm 1: Scaling algorithm for Theorem I.13

implies that

$$\left\| \mathrm{spec}\left(\rho_Y^{(i)}\right) - \mathrm{diag}\left(\boldsymbol{p}^{(i)}\right) \right\|_1 \leqslant \varepsilon$$

See full version.

To analyze our algorithm and prove Theorem I.13, we follow a three-step argument similar to the analysis in Section I-E. This has been used to great effect for operator scaling and tensor scaling in [4, 5, 52, 56] after identifying the appropriate potential function.

As we described in Section I-E, the appropriate potential functions to choose are the ones which are eigenvectors of an appropriate group action. In the matrix scaling case, we were acting by $\mathrm{T}(n) \times \mathrm{T}(n)$ and hence we chose the potential function to be permanent which is an eigenvector for this group action. In our algorithm, we are acting by the group corresponding to (direct products of) upper triangular matrices (this is known as the Borel subgroup). So for us, the right potential functions to consider are functions which are eigenvectors for the action of (tuples of) upper triangular matrices. One such class of functions are the so called highest weight vectors from representation theory[22], which we come to next.

What are highest weight vectors? We have the action of G on $V = \mathrm{Ten}(n_0; n_1, \ldots, n_d)$. Let us consider the space of degree k polynomial functions on V, denoted by $\mathbb{C}[V]_k$. The action of G on V induces an action of G on $\mathbb{C}[V]_k$ given by $(g \cdot P)(v) = P\left(g^{-1} \cdot v\right)$. Consider a tuple of vectors $\boldsymbol{\lambda} = \left(\boldsymbol{\lambda}^{(1)}, \ldots, \boldsymbol{\lambda}^{(d)}\right)$, $\boldsymbol{\lambda}^{(i)} \in \mathbb{Z}^{n_i}$. Then we say that P is a highest weight vector with weight $\boldsymbol{\lambda}$ if

$$g \cdot P = \prod_{i=1}^d \prod_{j=1}^{n_i} \left(g_{j,j}^{(i)}\right)^{\lambda_j^{(i)}} P$$

where $g = \left(g^{(1)}, \ldots, g^{(d)}\right)$ is such that $g^{(i)}$ is an upper triangular matrix for each i. Note that this necessitates $\sum_{j=1}^{n_i} \lambda_j^{(i)} = -k$ for each i. This also necessitates (not trivial to see why) that for all i, $\lambda_1^{(i)} \geqslant \cdots \geqslant \lambda_{n_i}^{(i)}$.

The following two properties of highest weight vectors will be crucial for our analysis:

1) [[17], see full version]: Let $\boldsymbol{p} \in P_+(n_1, \ldots, n_d)$ be a rational vector. Then $\boldsymbol{p} \in \Delta(X)$ iff there exists an integer $k \geqslant 1$ s.t. $\boldsymbol{\lambda} = k\boldsymbol{p}$ has integer entries and there exists a highest weight vector P with weight $\boldsymbol{\lambda}^*$ s.t. $(g \cdot P)(X) \neq 0$ for some $g \in G$. Here $\boldsymbol{\lambda}^* = \left(\left(-\lambda_{n_1}^{(1)}, \ldots, -\lambda_1^{(1)}\right), \ldots, \left(-\lambda_{n_d}^{(d)}, \ldots, -\lambda_1^{(d)}\right)\right)$. This extends a fact used in previous papers: the uniform vector is in $\Delta(X)$ iff some invariant polynomial does not vanish on X.

[22] Here we restrict our attention to the action on polynomials because that is what we need to describe the intuition for the analysis of the algorithm. But the discussion of weight vectors applies to arbitrary (rational) representations of the group G, see full version.

2) [see full version] The space of highest weight vectors with weight $\boldsymbol{\lambda}^*$ is spanned by polynomials P with integer coefficients that satisfy the following bound

$$|P(X)| \leqslant (n_1 \cdots n_d)^k \|X\|^k \tag{6}$$

This extends an identical bound in past papers from invariant polynomials to highest weight vectors.

We use classical constructions of highest weight vectors [4, 63, 64] to derive the second fact. These constructions are only semi-explicit (e.g., it is not clear if they can be evaluated efficiently), however they suffice for us because we only need a bound on their evaluations for their use as a potential function. We note that such bounds on their evaluations haven't been observed before in the invariant theory literature (except in [4] for the special case of invariants) whereas for us they are extremely crucial! We also emphasize that it is crucial for us that the bound is singly exponential in k. Some naive strategies of using solution sizes for linear systems only yield bounds that are doubly exponential in k.

The potential function we use is $\Phi(g) = |P(g \cdot X)|^{1/k}$. Here P is some highest weight vector of degree k (for some k), integer coefficients and weight λ^* that satisfies $(g \cdot P)(X) \neq 0$ as well as Eq. (6). Such a P exists by the discussion above. Using these properties, a three-step analysis, similar to the one in Section I-E, follows the following outline.

1) [**Lower bound**]: Since $(g \cdot P)(X)$ is nonzero for some g, we must have that $|P(g \cdot X)|$ is nonzero for a random choice of g. Furthermore, since we choose g to have integer coefficients, $|P(g \cdot X)| \geqslant 1$. After the normalization in Step (2), we get $\Phi(g) \geqslant 1/f(n_0, \ldots, n_d, d, b, M)$. It is not hard to see that $f(n_0, \ldots, n_d, b, M) \leqslant 2^b M^d (n_0 n_1 \cdots n_d)^2$.
2) [**Progress per step**]: $\Phi(g)$ increases at each step. Furthermore, if the current spectrum is "far" from the target spectrum, then one can explicitly bound the increase. Here the highest weight vector property of P, as well as Pinsker's inequality from information theory, play an important role.
3) [**Upper bound**]: $\Phi(g) \leqslant n_1 \cdots n_d$ always. This follows from Eq. (6) and the fact that we maintain the unit norm property of $g \cdot X$ after the normalization in Step (2) of the algorithm.

These three steps imply that in a polynomial number of iterations, one should get close to the target spectrum. A complete analysis is presented in the full version. Note that to ensure that we only use a polynomial amount of random bits for the initial randomization, we need the highest weight vectors to have degree at most exponential in the input parameters. This is achieved by relying on Derksen's degree bounds [65] (see full version).

G. Additional discussion

We would like to point out two important distinctions between the analysis for matrix scaling in Section I-E and our analysis here. First is that, as we have seen, there is a major difference between the uniform and the non-uniform versions of our problem - while this was not the case for matrix scaling. This phenomenon is general and is a distinction between commutative and non-commutative group actions. It has to do with the fact that all irreducible representations of commutative groups are one-dimensional, whereas for non-commutative groups they are not. Secondly, in the matrix scaling analysis, the upper bound was easy to obtain as well. Whereas for us, the upper bound step is the hardest and requires the use of deep results in representation theory. The upper bound steps were the cause of main difficulty in the papers [4, 5, 56] as well[23], and this is one key point of distinction between commutative and non-commutative group actions.

We believe that our framework of using the highest weight vectors as potential functions for the analysis of analytic algorithms is the right way to approach moment polytope problems - even beyond the cases that we consider in this paper.

The approach taken in [56] (for the case of $d = 2$) is one of reducing the non-uniform version of the problem to the uniform version (which was solved in [5] for the case of $d = 2$). The reduction is complicated and a bit ad hoc. We generalize this reduction to our setting ($d > 2$) (see full version), and provide a somewhat more principled view of the reduction along the way. However, it still seems rather specialized and mysterious compared to the general reduction in geometric invariant theory from the "non-uniform" to the "uniform" case (also known as the *shifting trick*, see full version).

[23]In some of the papers, the lower bound is the hard step, due to the use of a dual kind of potential function.

It is interesting to discuss some of the salient features and possible variations of Algorithm 1 (we expand on these points in the main text):

- **Iterations and randomness.** If we choose ℓ as the product of the denominators of all $p_j^{(i)}$, the algorithm terminates after at most $T = \mathrm{poly}\left(\frac{1}{\varepsilon}, d, \max_{i=0}^{d} n_i, b\right)$ iterations and uses $\log_2(M) = \mathrm{poly}(d, \max_{i=0}^{d} n_i, b)$ bits of randomness. For fixed or even inverse polynomial $\varepsilon > 0$, this is polynomial in the input size. In fact, this is better than the number of iterations in [56]: there, the number of iterations also depended on p_n^{-1} and q_n^{-1}. To get an algorithm with truly polynomial run time, one needs to truncate the group elements $g^{(i)}$'s up to polynomial number of bits after the decimal point. We provide an explanation on why this doesn't affect the performance of the algorithm in the full version.
- **Degenerate spectra.** If $\lambda^{(i)}$ is degenerate, i.e., $\lambda_j^{(i)} = \lambda_k^{(i)}$ for some $j \neq k$, then we may replace the Cholesky factors in step 3 by block upper triangular matrices, where the block sizes are the degeneracies - the set of such matrices is a so-called *parabolic subgroup* of the general linear group (see full version). Moreover, the random matrix g need only be generic up to action of the parabolic subgroup. In particular, when scaling to uniform spectra then no randomization is required and we can use Hermitian square roots, so Algorithm 1 reduces to the uniform tensor scaling algorithm of [4].
- **Singular spectra.** As written, Item 3 of Algorithm 1 fails if the spectra are singular, that is if for some i we have $r_i := \mathrm{rank}\,\mathrm{diag}(\boldsymbol{p}^{(i)}) < n_i$. However, in this case, one may first pass to a smaller tensor tensor X_+ obtained by restricting the i^{th} index to the last r_i coordinates. We show in the full version that X_+ is scalable by upper triangulars to marginals $\mathrm{diag}(p_{r_i}^{(i)}, \ldots, p_1^{(i)})$, $i \in [d]$ if and only if X is scalable by upper triangulars to $\mathrm{diag}(0, \ldots, 0, p_{r_i}^{(i)}, \ldots, p_1^{(i)})$, $i \in [d]$.

We discuss extensions of Algorithm 1 for more general varieties with "good" parametrizations in the full version.

H. Conclusions and open problems

We provide an efficient weak membership oracle for moment polytopes corresponding to a natural class of group actions on tensors. This generalizes recent works on operator and tensor scaling and also yields efficient algorithms for promise versions of the one-body quantum marginal problem and the asymptotic support of Kronecker coefficients. Our work leaves open several interesting questions some of which we state below.

- Improve the dependency on error ε in the running time of Algorithm 1 to $\mathrm{poly}(\log(1/\varepsilon))$. As discussed, this will immediately yield polynomial time algorithms for the one-body quantum marginal problem. This is open even for the uniform version of the problem. Here the notion of geodesic convexity of certain "capacity" optimization problems should play a key role (e.g. see [66]).
- Extend the weak membership oracle we develop to moment polytopes of other group actions, using Kirwan's gradient flow [22] as proposed in [21]. The quantitative tools developed in this paper naturally extend to this setup and we will elaborate on this in forthcoming work.
- Develop separation oracles for moment polytopes. A related question is: can we optimize over moment polytopes? This will have algorithmic applications on the problem of computing quantum functionals, as described in [28]. In this paper, Strassen's support functionals are generalized to quantum functionals, which are defined by convex optimization over the entanglement polytope. Thus, separation oracles for moment polytopes could lead to efficient algorithms for computing quantum functionals, which are important for comparing tensor powers (see [67–69]).
- Find natural instances of combinatorial optimization problems which can be encoded as moment polytopes. Some examples can be found in [1].

Acknowledgements

We would like to thank Shalev Ben David, Robin Kothari, Anand Natarajan, Frank Verstraete, John Watrous, John Wright, and Jeroen Zuiddam for interesting discussions.

References

[1] A. Garg, L. Gurvits, R. Oliveira, and A. Wigderson, "Algorithmic and optimization aspects of brascamp-lieb inequalities, via operator scaling," in *Proceedings of the Symposium on the*

Theory of Computing (STOC 2017), pp. 397–409, ACM, 2017, arXiv:1607.06711.

[2] A. Horn, "Doubly stochastic matrices and the diagonal of a rotation matrix," *American Journal of Mathematics*, vol. 76, no. 3, pp. 620–630, 1954.

[3] N. Linial, A. Samorodnitsky, and A. Wigderson, "A deterministic strongly polynomial algorithm for matrix scaling and approximate permanents," *STOC*, pp. 644–652, 1998.

[4] P. Bürgisser, A. Garg, R. Oliveira, M. Walter, and A. Wigderson, "Alternating minimization, scaling algorithms, and the null-cone problem from invariant theory," 2017, arXiv:1711.08039. Proceedings of Innovations in Theoretical Computer Science (ITCS 2018).

[5] A. Garg, L. Gurvits, R. Oliveira, and A. Wigderson, "A deterministic polynomial time algorithm for non-commutative rational identity testing," in *Proceedings of the Symposium on Foundations of Computer Science (FOCS 2016)*, pp. 109–117, IEEE, 2016, arXiv:1511.03730.

[6] L. Gurvits, "Combinatorial and algorithmic aspects of hyperbolic polynomials," 2005, arXiv:math/0404474.

[7] V. Guillemin and S. Sternberg, *Symplectic Techniques in Physics*. Cambridge University Press.

[8] A. C. da Silva, *Lectures on Symplectic Geometry*, vol. 1764 of *Lecture Notes in Mathematics*. Springer-Verlag, 2008.

[9] B. Kostant, "On convexity, the Weyl group and the Iwasawa decomposition," *Ann. scient. E.N.S*, vol. 6, pp. 413–455, 1973.

[10] M. F. Atiyah, "Convexity and commuting hamiltonians," *Bull. Lond. Math. Soc.*, vol. 14, pp. 1–15, 1982.

[11] V. Guillemin and S. Sternberg, "Convexity properties of the moment mapping," *Invent. Math.*, vol. 67, pp. 491–513, 1982.

[12] F. Kirwan, "Convexity properties of the moment mapping, III," *Inventiones mathematicae*, vol. 77, no. 3, pp. 547–552, 1984.

[13] J. M. Landsberg, *Tensors: geometry and applications*, vol. 128. AMS, 2012.

[14] W. Hackbusch, *Tensor spaces and numerical tensor calculus*, vol. 42. Springer, 2012.

[15] A. A. Klyachko, "Quantum marginal problem and N-representability," in *Journal of Physics: Conference Series*, vol. 36, p. 72, 2006, arXiv:quant-ph/0511102.

[16] Y.-K. Liu, "Consistency of local density matrices is QMA-complete," *RANDOM*, 2006.

[17] L. Ness and D. Mumford, "A stratification of the null cone via the moment map," *American Journal of Mathematics*, vol. 106, no. 6, pp. 1281–1329, 1984.

[18] M. Brion, "Sur l'image de l'application moment," in *Séminaire d'algebre Paul Dubreil et Marie-Paule Malliavin*, vol. 1296 of *Lecture Notes in Mathematics*, pp. 177–192, Springer, 1987.

[19] N. Ressayre, "Geometric invariant theory and the generalized eigenvalue problem," *Inventiones mathematicae*, vol. 180, no. 2, pp. 389–441, 2010, arXiv:0903.1187.

[20] M. Vergne and M. Walter, "Inequalities for moment cones of finite-dimensional representations," *Journal of Symplectic Geometry*, vol. 15, no. 4, pp. 1209–1250, 2017, arXiv:1410.8144.

[21] M. Walter, *Multipartite Quantum States and their Marginals*. PhD thesis, ETH Zurich, 2014, arXiv:1410.6820.

[22] F. C. Kirwan, *Cohomology of quotients in symplectic and algebraic geometry*, vol. 31. Princeton University Press, 1984.

[23] M. Walter, B. Doran, D. Gross, and M. Christandl, "Entanglement polytopes: multiparticle entanglement from single-particle information," *Science*, vol. 340, no. 6137, pp. 1205–1208, 2013, arXiv:1208.0365.

[24] A. Sawicki, M. Oszmaniec, and M. Kuś, "Convexity of momentum map, Morse index, and quantum entanglement," *Reviews in Mathematical Physics*, vol. 26, no. 03, p. 1450004, 2014, arXiv:1208.0556.

[25] P. Bürgisser, J. M. Landsberg, L. Manivel, and J. Weyman, "An overview of mathematical issues arising in the geometric complexity theory approach to VP≠VNP," *SIAM Journal on Computing*, vol. 40, no. 4, pp. 1179–1209, 2011, arXiv:0907.2850.

[26] P. Bürgisser and C. Ikenmeyer, "Geometric complexity theory and tensor rank," in *Proceedings of the Symposium on the Theory of Computing (STOC 2011)*, pp. 509–518, ACM, 2011, arXiv:1011.1350.

[27] M. Christandl, B. Doran, and M. Walter, "Computing multiplicities of lie group representations," in *Proceedings of the Symposium on Foundations of Computer Science (FOCS 2012)*, pp. 639–648, IEEE, 2012, arXiv:1204.4379.

[28] M. Christandl, P. Vrana, and J. Zuiddam, "Universal points in the asymptotic spectrum of tensors," 2017, arXiv:1709.07851.

[29] C. H. Bennett, S. Popescu, D. Rohrlich, J. A. Smolin, and A. V. Thapliyal, "Exact and asymptotic measures of multipartite pure-state entanglement," *Physical Review A*, vol. 63, no. 1, p. 012307, 2000.
[30] W. Fulton and J. Harris, *Representation theory: a first course*, vol. 129. Springer, 2013.
[31] R. P. Stanley, "Positivity problems and conjectures in algebraic combinatorics," *Mathematics: frontiers and perspectives*, vol. 295, p. 319, 2000.
[32] K. D. Mulmuley, "Geometric complexity theory VI: The flip via positivity," 2007, arXiv:0704.0229.
[33] M. Christandl and G. Mitchison, "The spectra of quantum states and the Kronecker coefficients of the symmetric group," *Communications in Mathematical Physics*, vol. 261, no. 3, pp. 789–797, 2006, arXiv:quant-ph/0409016.
[34] M. Christandl, A. W. Harrow, and G. Mitchison, "Nonzero Kronecker coefficients and what they tell us about spectra," *Communications in Mathematical Physics*, vol. 270, no. 3, pp. 575–585, 2007, arXiv:quant-ph/0511029.
[35] C. Ikenmeyer, K. D. Mulmuley, and M. Walter, "On vanishing of Kronecker coefficients," *computational complexity*, vol. 26, no. 4, pp. 949–992, 2017, arXiv:1507.02955.
[36] A. Knutson and T. Tao, "The honeycomb model of gln(c) tensor products i: Proof of the saturation conjecture," *Journal of the American Mathematical Society*, vol. 12, no. 4, pp. 1055–1090, 1999.
[37] P. Bürgisser and C. Ikenmeyer, "Deciding positivity of Littlewood-Richardson coefficients," *SIAM J. Discrete Math.*, vol. 27, no. 4, pp. 1639–1681, 2012.
[38] M. F. Atiyah, "Convexity and commuting Hamiltonians," *Bulletin of the London Mathematical Society*, vol. 14, no. 1, pp. 1–15, 1982.
[39] V. Guillemin and S. Sternberg, "Convexity properties of the moment mapping," *Inventiones mathematicae*, vol. 67, no. 3, pp. 491–513, 1982.
[40] V. Guillemin and S. Sternberg, "Convexity properties of the moment mapping. II," *Inventiones mathematicae*, vol. 77, no. 3, pp. 533–546, 1984.
[41] M. Brion, "On the general faces of the moment polytope," *International Mathematics Research Notices*, vol. 1999, no. 4, pp. 185–201, 1999.
[42] A. Berenstein and R. Sjamaar, "Coadjoint orbits, moment polytopes, and the Hilbert–Mumford criterion," *Journal of the American Mathematical Society*, vol. 13, no. 2, pp. 433–466, 2000, arXiv:math/9810125.
[43] P. Belkale, "The tangent space to an enumerative problem," in *Proceedings of the International Congress of Mathematicians (ICM 2010)*, vol. 2, pp. 405–426, 2010.
[44] A. Klyachko, "Quantum marginal problem and representations of the symmetric group," *quant-ph/0409113*, 2004.
[45] S. Daftuar and P. Hayden, "Quantum state transformations and the Schubert calculus," *Annals of Physics*, vol. 315, no. 1, pp. 80–122, 2005, arXiv:quant-ph/0410052.
[46] M. Altunbulak and A. Klyachko, "The Pauli principle revisited," *Communications in Mathematical Physics*, vol. 282, no. 2, pp. 287–322, 2008, arXiv:0802.0918.
[47] A. Higuchi, A. Sudbery, and J. Szulc, "One-qubit reduced states of a pure many-qubit state: polygon inequalities," *Physical Review Letters*, vol. 90, no. 10, p. 107902, 2003, arXiv:quant-ph/0209085.
[48] S. Bravyi, "Requirements for compatibility between local and multipartite quantum states," *Quantum Information and Computation*, vol. 4, pp. 12–26, 2003, arXiv:quant-ph/0301014.
[49] M. Franz, "Moment polytopes of projective G-varieties and tensor products of symmetric group representations," *Journal of Lie Theory*, vol. 12, no. 2, pp. 539–549, 2002.
[50] B. Sturmfels, *Algorithms in invariant theory*. Springer, 2008.
[51] H. Derksen and G. Kemper, *Computational invariant theory*. Springer, 2015.
[52] L. Gurvits, "Classical complexity and quantum entanglement," *Journal of Computer and System Sciences*, vol. 69, no. 3, pp. 448–484, 2004.
[53] G. Ivanyos, Y. Qiao, and K. Subrahmanyam, "Constructive non-commutative rank computation is in deterministic polynomial time," no. 5, 2017, arXiv:1512.03531. Proceedings of Innovations in Theoretical Computer Science (ITCS 2017).
[54] H. Derksen and V. Makam, "Polynomial degree bounds for matrix semi-invariants," 2015, arXiv:1512.03393.
[55] F. Verstraete, J. Dehaene, and B. De Moor, "Normal forms and entanglement measures for multipartite quantum states," *Physical Review*

A, vol. 68, no. 1, p. 012103, 2003, arXiv:quant-ph/0105090.

[56] C. Franks, "Operator scaling with specified marginals," *STOC*, 2018, arXiv:1801.01412.

[57] R. Sinkhorn, "A relationship between arbitrary positive matrices and doubly stochastic matrices," *The Annals of Mathematical Statistics*, vol. 35, pp. 876–879, 1964.

[58] R. Sinkhorn and P. Knopp, "Concerning nonnegative matrices and doubly stochastic matrices," *Pacific Journal of Mathematics*, vol. 21, no. 2, pp. 343–348, 1967.

[59] P. Bürgisser, M. Christandl, K. D. Mulmuley, and M. Walter, "Membership in moment polytopes is in NP and coNP," *SIAM Journal on Computing*, vol. 46, no. 3, pp. 972–991, 2017, arXiv:1511.03675.

[60] R. Orús, "A practical introduction to tensor networks: Matrix product states and projected entangled pair states," *Annals of Physics*, vol. 349, pp. 117–158, 2014, arXiv:1306.2164.

[61] P. Bürgisser, M. Clausen, and M. A. Shokrollahi, *Algebraic complexity theory*, vol. 315. Springer Science & Business Media, 2013.

[62] L. Gurvits and P. N. Yianilos, "The deflation-inflation method for certain semidefinite programming and maximum determinant completion problems," *Technical Report, NECI*, 1998.

[63] C. Procesi, "Lie groups," 2007.

[64] P. Bürgisser and C. Ikenmeyer, "Explicit lower bounds via geometric complexity theory," in *Proceedings of the Symposium on the Theory of Computing (STOC 2013)*, pp. 141–150, ACM, 2013, arXiv:1210.8368.

[65] H. Derksen, "Polynomial bounds for rings of invariants," *Proceedings of the American Mathematical Society*, vol. 129, no. 4, pp. 955–963, 2001.

[66] Z. Allen-Zhu, A. Garg, Y. Li, R. Oliveira, and A. Wigderson, "Operator scaling via geodesically convex optimization, invariant theory and polynomial identity testing," *STOC*, 2018.

[67] V. Strassen, "The asymptotic spectrum of tensors and the exponent of matrix multiplication," in *IEEE 27th Annual Symposium on Foundations of Computer Science (FOCS)*, pp. 49–54, IEEE, 1986.

[68] V. Strassen, "The asymptotic spectrum of tensors.," *Journal für die reine und angewandte Mathematik*, vol. 384, pp. 102–152, 1988.

[69] V. Strassen, "Degeneration and complexity of bilinear maps: some asymptotic spectra," *J. Reine Angew. Math.*, vol. 413, pp. 127–180, 1991.