Optimization, Complexity and Invariant Theory

Peter Bürgisser
Institut für Mathematik, Technische Universität Berlin, Germany

Abstract

Invariant and representation theory studies symmetries by means of group actions and is a well established source of unifying principles in mathematics and physics. Recent research suggests its relevance for complexity and optimization through quantitative and algorithmic questions. The goal of the talk is to give an introduction to new algorithmic and analysis techniques that extend convex optimization from the classical Euclidean setting to a general geodesic setting. We also point out surprising connections to a diverse set of problems in different areas of mathematics, statistics, computer science, and physics.

2012 ACM Subject Classification Theory of computation → Continuous optimization

Keywords and phrases geometric invariant theory, geodesic optimization, non-commutative optimization, null cone, operator scaling, moment polytope, orbit closure intersection, geometric programming

Digital Object Identifier 10.4230/LIPIcs.STACS.2021.1

Category Invited Talk

Related Version *Full Version*: https://arxiv.org/abs/1910.12375

Funding *Peter Bürgisser*: Supported by the ERC under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 787840.)

Acknowledgements The talk (and this write-up) are mainly based on the joint articles [14, 15] with Cole Franks, Ankit Garg, Rafael Oliveira, Michael Walter and Avi Wigderson. I am grateful to my coauthors for inspiration and enlightening discussions and thank Levent Dogan, Yinan Li, Visu Makam, Harold Nieuwboer and Philipp Reichenbach for valuable feedback.

1 Introduction

Consider a group G that acts by linear transformations on the complex Euclidean space $V = \mathbb{C}^m$. This partitions V into *orbits*: For a vector $v \in V$, the orbit $\mathcal{O}_v$ consists of all vectors of the form $g \cdot v$ to which the action of a group element $g \in G$ can map v.

The most basic algorithmic question in this setting is as follows. Given a vector $v \in V$, compute (or approximate) the smallest ℓ_2-norm of any vector in the orbit of v, that is, $\inf\{\|w\|_2 : w \in \mathcal{O}_v\}$. Remarkably, this simple question, for different groups and actions, captures natural important problems in computational complexity, algebra, analysis, statistics and quantum information. When restricted to commutative groups G, this amount to unconstrained geometric programming (see Section 2). In particular, this already captures all linear programming problems!

Starting with [37, 39], a series of recent works including [38, 16, 34, 62, 2, 13] designed algorithms and analysis tools to handle this basic and other related optimization problems over *non-commutative* groups G. In all these works, the groups at hand are products of at least two copies of rather specific linear groups ($\mathrm{SL}(n)$'s or tori), to support the algorithms and analysis. These provided efficient solutions for some applications, and *through algorithms*, the resolution of some purely structural mathematical open problems.

We mention here some of the diverse applications of the paradigm of optimization over non-commutative groups.

38th International Symposium on Theoretical Aspects of Computer Science (STACS 2021).
Editors: Markus Bläser and Benjamin Monmege; Article No. 1; pp. 1:1–1:20
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1. **Algebraic identities:** Given an arithmetic formula (with inversion gates) in non-commuting variables, is it identically zero?
2. **Quantum information:** Given density matrices describing local quantum states of various parties, is there a global pure state consistent with the local states?
3. **Eigenvalues of sums of Hermitian matrices:** Given three real n-vectors, do there exist three Hermitian $n \times n$ matrices A, B, C with these *prescribed* spectra, such that $A + B = C$?
4. **Analytic inequalities:** Given m linear maps $A_i : \mathbb{R}^n \to \mathbb{R}^{n_i}$ and $p_1, \ldots, p_m \geq 0$, does there exist a finite constant C such that for all integrable functions $f_i : \mathbb{R}^{n_i} \to \mathbb{R}_+$ we have

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

   These inequalities are the celebrated Brascamp-Lieb inequalities, which capture the Cauchy-Schwarz, Hölder, Loomis-Whitney, and many further inequalities.
5. **Maximum Likelihood Estimation** Consider a centered Gaussian random variable $Y \in \mathbb{R}^n$ with a covariance matrix Ψ being an element of the matrix normal model $\mathcal{M}(p,q) = \{\Psi_1 \otimes \Psi_2 \mid \Psi_1 \in \mathrm{PD}_p, \Psi_2 \in \mathrm{PD}_q\}$. What is the number of samples needed to achieve almost surely the existence and uniqueness of maximum likelihood estimators?

At first glance, it is far from obvious that solving any of these problems has any relation to either optimization or groups. However, not only symmetries naturally exist in all of them, but they also help both in formulating them as optimization problems over groups, suggesting natural algorithms (or at least heuristics) for them, and finally in providing tools for analyzing these algorithms. It perhaps should be stressed that symmetries exist in many examples in which the relevant groups are commutative (e.g., perfect matching in bipartite graphs, matrix scaling, and more generally in linear, geometric, and hyperbolic programming); however in these cases, standard convex optimization or combinatorial algorithms can be designed and analyzed without any reference to these existing symmetries.

Polynomial time algorithms were first given in [37] for Problem 1 (the works [50, 27, 49] later discovered completely different algebraic algorithms), in [13] for Problem 2 (cf. [79] and the structural results [57, 25, 22, 21, 81, 80, 19]), in [59, 64, 68, 17, 34] for Problem 3 (the celebrated structural result in [59] and the algorithmic results of [64, 68, 17] solved the decision problem, while [34] solved the search problem), and in [38] for Problem 4. However the algorithms in [38, 34, 13] remain exponential time in various input parameters, exemplifying only one aspect of many in which the current theory and understanding is lacking. Problem 5 was recently solved by in [4, 29], proving a conjecture in [32] and generalized to tensor normal models in [30].

The unexpected connections revealed in this study are far richer than the mere relevance of optimization and symmetries to such problems. One type are connections between problems in disparate fields. For example, the analytic Problem 4 turns out to be a special case of the algebraic Problem 1. Moreover, Problem 1 has (well-studied) differently looking but equivalent formulations in quantum information theory and in invariant theory, which are automatically solved by the same algorithm. Another type of connections are of purely structural open problems solved through such geodesic algorithms, reasserting the importance of the computational lens in mathematics. One was the first dimension-independent bound on the Paulsen problem in operator theory, obtained ingeniously through such an algorithm in [62] (this work was followed by [46], who gave a strikingly simpler proof and stronger bounds). Another was a quantitative bound on the continuity of the best constant C in

Problem 4 (in terms of the input data), important for non-linear variants of such inequalities. This bound was obtained through the algorithm in [38] and relies on its efficiency; previous methods used compactness arguments that provided no bounds.

We have no doubt that more unexpected applications and connections will follow. The most extreme and speculative perhaps among such potential applications is to develop a deterministic polynomial-time algorithm for the polynomial identity testing (PIT) problem. Such an algorithm will imply major algebraic or Boolean lower bounds, namely either separating VP from VNP, or proving that NEXP has no small Boolean circuits [51]. We note that this goal was a central motivation of the initial work in this sequence [37], which provided such a deterministic algorithm for Problem 1 above, the non-commutative analog of PIT. The "real" PIT problem (in which variables commute) also has a natural group of symmetries acting on it, which does not quite fall into the frameworks developed so far. Yet, the hope of proving lower bounds via optimization methods is a fascinating (and possibly achievable) one. This agenda of hoping to shed light on the PIT problem by the study of invariant theoretic questions was formulated in the fifth paper of the Geometric Complexity Theory (GCT) series [66, 67], but see [40].

In this talk, we describe the main results of the paper [15], which unifies and generalizes the above mentioned works. A key to all of them are the notions of *geodesic convexity* (which generalizes the familiar Euclidean notion of convexity) and the *moment map* (which generalizes the familiar Euclidean gradient) in the curved space and new metrics induced by the group action. The paper [15] naturally extends the familiar first and second order methods of standard convex optimization. Geodesic analogs of these methods are designed, which, respectively, have oracle access to first and second order "derivatives" of the function being optimized, and apply to any (reductive) group action. The first order method developed (which is a non-commutative version of gradient descent) replaces and extends the use of "alternate minimization" in most past works, and thus can accommodate more general group actions. For instance, this covers the cases of symmetric tensors (bosons) and antisymmetric tensors (fermions) with the standard action of $\mathrm{SL}(n)$, where alternating minimization does not apply. The second order method developed in [15] greatly generalizes the one used for the particular group action corresponding to operator scaling in [2]. It may be thought of as a geodesic analog of the "trust region method" [24] or the "box-constrained Newton method" [23, 3] applied to a regularized function. For both methods, in this non-commutative setting, we recover the familiar convergence behavior of the classical commutative case: to achieve "proximity" ε to the optimum, the first order method converges in $O(1/\varepsilon)$ iterations and the second order method in $O(\operatorname{poly}\log(1/\varepsilon))$ iterations.

As in the commutative case, the fundamental challenge is to understand the "constants" hidden in the big-O notation of each method. These depend on "smoothness" properties of the function optimized, which in turn are determined by the action of the group G on the space V that defines the optimization problem. The main technical contributions of the theory developed in [15] is to identify the key parameters which control this dependence, and to bound them for various actions to obtain concrete running time bounds. These parameters depend on a combination of algebraic and geometric properties of the group action, in particular the irreducible representations occurring in it. As mentioned, despite the technical complexity of defining (and bounding) these parameters, the way they control convergence of the algorithms is surprisingly elegant. The paper [15] also develops important technical tools which naturally extend ones common in the commutative theory, including regularizers, diameter bounds, numerical stability, and initial starting points, which are key to the design and analysis of the presented (and hopefully future) algorithms in the geodesic setting. As in previous works, we also address other optimization problems beyond the basic "norm minimization" question above, in particular the minimization of the moment map, and the membership problem for *moment polytopes*; a very rich class of polytopes (typically with exponentially many vertices and facets) which arises magically from any such group action.

2 Non-commutative optimization

We now give an introduction to non-commutative optimization and contrast its geometric structure and convexity properties with the familiar commutative setting. The basic setting is that of a continuous group G acting linearly on an m-dimensional complex vector space $V \cong \mathbb{C}^m$. Think of G as either the group of $n \times n$ complex invertible matrices, denoted $\mathrm{GL}(n)$, or the group of diagonal such matrices, denoted $\mathrm{T}(n)$.[1] The latter corresponds to the commutative case and the former is a paradigmatic example of the non-commutative case. An (linear) *action* (also called *representation*) of a group G on a complex vector space V is a group homomorphism $\pi : G \to \mathrm{GL}(V)$, that is, an association of an invertible linear map $\pi(g) \colon V \to V$ for every group element $g \in G$ satisfying $\pi(g_1 g_2) = \pi(g_1)\pi(g_2)$ for all $g_1, g_2 \in G$.[2] Further suppose that V is also equipped with a Hermitian inner product $\langle \cdot, \cdot \rangle$ and hence a norm $\|v\| := \langle v, v \rangle$.

Given a vector $v \in V$ one can consider the optimization problem of taking the infimum of the norm in the orbit of the vector v under the action of G. More formally, we define the *capacity* of v (with respect to π) by

$$\mathrm{cap}(v) := \inf_{g \in G} \|\pi(g) v\|.$$

This notion generalizes the matrix and operator capacities developed in [45, 42] (to see this, carry out the optimization over one of the two group variables) as well as the polynomial capacity of [44]. It turns out that this simple-looking optimization problem is already very general in the commutative case and, in the non-commutative case, captures all examples discussed in the introduction.

Let us first consider the commutative case, $G = \mathrm{T}(n)$ acting on V. In this simple case, all actions π have a very simple form. We give two equivalent descriptions, first of how any representation π splits into one-dimensional irreducible representations, and another describing π as a natural scaling action on n-variate polynomials with m monomials.

The irreducible representations are given by an orthonormal basis $v_1, \ldots, v_m$ of V such that the v_j are simultaneous eigenvectors of all the matrices $\pi(g)$. That is, for all $g = \mathrm{diag}(g_1, \ldots, g_n) \in \mathrm{T}(n)$ and $j \in [m]$, we have

$$\pi(g) v_j = \lambda_j(g) v_j, \quad \text{where} \quad \lambda_j(g) = \prod_{i=1}^n g_i^{\omega_{j,i}} \tag{2.1}$$

for fixed integer vectors $\omega_1, \ldots, \omega_m \in \mathbb{Z}^n$, which are called *weights* and encode the simultaneous eigenvalues, and completely determine the action. Below we also refer to the weights of a representation π of $\mathrm{GL}(n)$, defined as the weights of π restricted to $\mathrm{T}(n)$.

A natural way to view all these actions is as follows. The natural action of $\mathrm{T}(n)$ on $\mathbb{C}^n$ by matrix-vector multiplication induces an action of $\mathrm{T}(n)$ on n-variate polynomials $V = \mathbb{C}[x_1, x_2, \ldots, x_n]$: simply, any group element $g = \mathrm{diag}(g_1, \ldots, g_n)$ "scales" each x_i to $g_i x_i$. Note that any monomial $x^\omega = \prod_{i=1}^n x_i^{\omega(i)}$ (where ω is the integer vector of exponents) is an eigenvector of this action, with an eigenvalue $\lambda(g) = \prod_{i=1}^n g_i^{\omega(i)}$.

[1] The theory works whenever the group is connected, algebraic and reductive.
[2] We further assume that π is a morphism of algebraic groups, i.e., given by rational functions.

Now fix m integer vectors ω_j as above. Consider the linear space of n-variate Laurent polynomials (monomials may have negative exponents) with the following m monomials: $v_j = x^{\omega_j} = \prod_{i=1}^n x_i^{\omega_{j,i}}$. The action on any polynomial $v = \sum_{j=1}^m c_j v_j$ is precisely the one described above, scaling each coefficient by the eigenvalue of its monomial. The norm $\|v\|$ of a polynomial is the sum of the square moduli of its coefficients. Now let us calculate the capacity of this action. For any $v = \sum_{j=1}^m c_j v_j$,

$$\mathrm{cap}(v)^2 = \inf_{g_1,\dots,g_n \in \mathbb{C}^*} \sum_{j=1}^m |c_j|^2 \prod_{i=1}^n |g_i|^{2\omega_{j,i}} = \inf_{x\in\mathbb{R}^n} \sum_{j=1}^m |c_j|^2 e^{x\cdot\omega_j}, \tag{2.2}$$

where we used the change of variables $x_i = \log|g_i|^2$, which makes the problem convex (in fact, log-convex). This class of optimization problems (of optimizing norm in the orbit of a commutative group) is known as *geometric programming* and is well-studied in the optimization literature (see, e.g., Chapter 4.5 in [10]). Hence for non-commutative groups, one can view computing $\mathrm{cap}(v)$ as *non-commutative geometric programming*. Is there a similar change of variables that makes the problem convex in the non-commutative case? It does not seem so. However, the non-commutative case also satisfies a notion of convexity, known as geodesic convexity, which we will study next.

2.1 Geodesic convexity

Geodesic convexity generalizes the notion of convexity in the Euclidean space to arbitrary Riemannian manifolds. We will not go into the notion of geodesic convexity in this generality but just mention what it amounts to in our concrete setting of norm optimization for $G = \mathrm{GL}(n)$.

It turns out the appropriate way to define geodesic convexity in this case is as follows. Fix an action π of $\mathrm{GL}(n)$ and a vector v. Then $\log\|\pi(e^{tH}g)v\|$ is convex in the real parameter t for every Hermitian matrix H and $g \in \mathrm{GL}(n)$. This notion of convexity is quite similar to the notion of Euclidean convexity, where a function is convex iff it is convex along all lines. However, it is far from obvious how to import optimization techniques from the Euclidean setting to work in this non-commutative geodesic setting. An essential ingredient we describe next is the geodesic notion of a gradient, called the *moment map*.

2.2 Moment map

The moment map is by definition the gradient of the function $\log\|\pi(g)v\|$ (understood as a function of v), at the identity element of the group, $g = I$. It captures how the norm of the vector v changes when we act on it by infinitesimal perturbations of the identity.

Again, we start with the commutative case $G = \mathrm{T}(n)$ acting on the multivariate Laurent polynomials. For a direction vector $h \in \mathbb{R}^n$ and a real perturbation parameter t, let $e^{th} = \mathrm{diag}\left(e^{th_1}, \dots, e^{th_n}\right)$. Then, for $G = \mathrm{T}(n)$, the moment map is the function $\mu\colon V \setminus \{0\} \to \mathbb{R}^n$, defined by the following property:

$$\mu(v)\cdot h = \partial_{t=0}\left[\log\left\|\pi(\mathrm{diag}(e^{th})v\right\|\right],$$

for all $h \in \mathbb{R}^n$. That is, the directional derivative in direction h is given by the dot product $\mu(v)\cdot h$. Here one can see that the moment map matches the notion of Euclidean gradient. For the action of $\mathrm{T}(n)$ in Equation (2.1), we have

$$\mu(v) = \nabla_{x=0}\log\left(\sum_{j=1}^m |c_j|^2 e^{x\cdot\omega_j}\right) = \frac{\sum_{j=1}^m |c_j|^2\omega_j}{\sum_{j=1}^m |c_j|^2}. \tag{2.3}$$

Note that the gradient $\mu(v)$ at any point v is a convex combination of the weights. Viewing v as a polynomial, the gradient thus belongs to the so-called *Newton polytope* of v, namely the convex hull of the exponent vectors of its monomials. Conversely, every point in that polytope is a gradient of some polynomial v with these monomials. We will soon return to this curious fact!

We now proceed to the non-commutative case, focusing on $G = \mathrm{GL}(n)$. Denote by $\mathrm{Herm}(n)$ the set of $n \times n$ complex Hermitian matrices.[3] Here directions will be parametrized by $H \in \mathrm{Herm}(n)$. For the case of $G = \mathrm{GL}(n)$, the moment map is the function $\mu\colon V \setminus \{0\} \to \mathrm{Herm}(n)$ defined (in complete analogy to the commutative case above) by the following property that

$$\mathrm{tr}[\mu(v)H] = \partial_{t=0}\left[\log \left\|\pi(e^{tH})v\right\|\right]$$

for all $H \in \mathrm{Herm}(n)$. That is, the directional derivative in direction H is given by $\mathrm{tr}[\mu(v)H]$.

In the commutative case, Equation (2.3) is a convex combination of the weights ω_j. Thus, the image of μ is the convex hull of the weights – a convex polytope. This brings us to moment polytopes.

2.3 Moment polytopes

One may ask whether the above fact is true for actions of $\mathrm{GL}(n)$: is the set $\{\mu(v) : v \in V\setminus\{0\}\}$ convex? This turns out to be blatantly false: for instance, for the action of $\mathrm{GL}(n)$ on $\mathbb{C}^n$ by matrix-vector multiplication the moment map is $\mu(v) = vv^\dagger/\|v\|^2$, and its image is clearly not convex. However, there is still something deep and non-trivial that can be said. Given a Hermitian matrix $H \in \mathrm{Herm}(n)$, define its *spectrum* to be the vector of its eigenvalues arranged in non-increasing order. That is, $\mathrm{spec}(H) := (\lambda_1, \ldots, \lambda_n)$, where $\lambda_1 \geq \cdots \geq \lambda_n$ are the eigenvalues of H. Amazingly, the set of spectra of moment map images, that is,

$$\Delta := \left\{\mathrm{spec}\big(\mu(v)\big) : 0 \neq v \in V\right\}, \tag{2.4}$$

is a convex polytope for every representation π [70, 60, 5, 41, 55]! These polytopes are called *moment polytopes*.

Let us mention two important examples of moment polytopes. The examples are for actions of products of $\mathrm{GL}(n)$'s but the above definitions generalize almost immediately.

▶ **Example 2.1** (Horn's problem). Let $G = \mathrm{GL}(n) \times \mathrm{GL}(n) \times \mathrm{GL}(n)$ act on $V = \mathrm{Mat}(n) \oplus \mathrm{Mat}(n)$ as follows: $\pi(g_1, g_2, g_3)(X, Y) := (g_1 X g_3^{-1}, g_2 Y g_3^{-1})$. The moment map in this case is

$$\mu(X, Y) = \frac{(XX^\dagger, YY^\dagger, -(X^\dagger X + Y^\dagger Y))}{\|X\|_F^2 + \|Y\|_F^2}.$$

Using that $XX^\dagger$ and $X^\dagger X$ are positive semidefinite and isospectral, we obtain the following moment polytopes, which characterize the eigenvalues of sums of Hermitian matrices, i.e., Horn's problem (see, e.g., [36, 9]):

$$\Delta = \left\{(\mathrm{spec}(A), \mathrm{spec}(B), \mathrm{spec}(-A-B)) \mid A, B \in \mathrm{Mat}(n),\ A \geq 0,\ B \geq 0,\ \mathrm{tr}\,A + \mathrm{tr}\,B = 1\right\}.$$

These polytopes are known as the *Horn polytopes* and correspond to Problem 3 in the introduction. They have been characterized mathematically in [58, 59, 7, 72] and algorithmically in [64, 68, 17].

[3] The reason we are restricting to directions in $\mathbb{R}^n$ in the $\mathrm{T}(n)$ case and to directions in Herm_n in the $\mathrm{GL}(n)$ case is that imaginary and skew-Hermitian directions, respectively, do not change the norm.

The preceding is one of the simplest example of a moment polytope associated with the representation of a quiver (the star quiver with two edges); see [31] for this notion. Quiver representations are relevant for the solution of Problem 5.

▶ **Example 2.2** (Tensor action). $G = \mathrm{GL}(n_1) \times \mathrm{GL}(n_2) \times \mathrm{GL}(n_3)$ acts on $V = \mathbb{C}^{n_1} \otimes \mathbb{C}^{n_2} \otimes \mathbb{C}^{n_3}$ as follows: $\pi(g_1, g_2, g_3)v := (g_1 \otimes g_2 \otimes g_3)v$. We can think of vectors $v \in V$ as tripartite quantum states with local dimensions n_1, n_2, n_3. Then the moment map for this group action captures precisely the notion of *quantum marginals*. That is, $\mu(v) = (\rho_1, \rho_2, \rho_3)$, where $\rho_k = \mathrm{tr}_{k^c}(vv^\dagger)$ denotes the reduced density matrix describing the state of the k^{th} particle. This corresponds to Problem 2 in the introduction.

The moment polytopes in this case are known as *Kronecker polytopes*, since they can be equivalently described in terms of the Kronecker coefficients of the symmetric group. These polytopes have been studied in [57, 25, 22, 21, 81, 80, 19, 78, 13].

There is a more refined notion of a moment polytope. One can look at the collection of spectra of moment maps of vectors in the orbit of a particular vector $v \in V$. Surprisingly, its closure,

$$\Delta(v) := \overline{\left\{\mathrm{spec}\big(\mu(w)\big) : w \in \mathcal{O}_v\right\}}$$

is a convex polytope as well, called the *moment polytope of* v [70, 11]! It can equivalently be defined as the spectra of moment map images of the orbit's closure in projective space.

2.4 Null cone

Fix a representation π of the group G on a vector space V (again assume $G = \mathrm{T}(n)$ or $G = \mathrm{GL}(n)$ for simplicity). The *null cone* for this group action is defined as the set of vectors v such that $\mathrm{cap}(v) = 0$:

$$\mathcal{N} := \{v \in V : \mathrm{cap}(v) = 0\}.$$

In other words, v is in the null cone if and only if 0 lies in the orbit-closure of v. It is of importance in invariant theory due to the results of Hilbert and Mumford [47, 69] which state that the null cone is the algebraic variety defined by non-constant homogeneous invariant polynomials of the group action (see, e.g., the excellent textbooks [26, 76]).

Let us see what the null cone for the action of $\mathrm{T}(n)$ in Equation (2.1) is. Recall from Equation (2.2), the formulation for $\mathrm{cap}(v)$. It is easy to see that $\mathrm{cap}(v) = 0$ iff there exists $x \in \mathbb{R}^n$ such that $x \cdot \omega_j < 0$ for all $j \in \mathrm{supp}(v)$, where $\mathrm{supp}(v) = \{j \in [m] : c_j \neq 0\}$ for $v = \sum_{j=1}^m c_j v_j$. Thus the property of v being in the null cone is captured by a simple linear program defined by $\mathrm{supp}(v)$ and the weights ω_j's. Hence the null cone membership problem for non-commutative group actions can be thought of as *non-commutative linear programming*.

We know by Farkas' lemma that there exists $x \in \mathbb{R}^n$ such that $x \cdot \omega_j < 0$ for all $j \in \mathrm{supp}(v)$ iff 0 does not lie in $\mathrm{conv}\{\omega_j : j \in \mathrm{supp}(v)\}$. In other words, $\mathrm{cap}(v) = 0$ iff $0 \notin \Delta(v)$. Is this true in the non-commutative world? It is! This is the Kempf-Ness theorem [53] and it is a consequence of the geodesic convexity of the function $g \to \log\|\pi(g)v\|$. The Kempf-Ness theorem can be thought of as a *non-commutative duality theory* paralleling the linear programming duality given by Farkas' lemma (which corresponds to the commutative world). Let us now mention an example of an interesting null cone in the non-commutative case.

▶ **Example 2.3** (Operator scaling, or left-right action). The action of the group $G = \mathrm{SL}(n) \times \mathrm{SL}(n)$ on $\mathrm{Mat}(n)^k$ via $\pi(g,h)(X_1, \ldots, X_k) := (gX_1h^T, \ldots, gX_kh^T)$ is called the *left-right* action. (Recall $\mathrm{SL}(n)$ denotes the group of $n \times n$ matrices with determinant 1.) The null cone for this action captures *non-commutative singularity* (see, e.g., [50, 37, 27, 49]) and Problem 1 in the introduction. The left-right action has been crucial in getting deterministic polynomial time algorithms for the non-commutative rational identity testing problem [50, 37, 27, 49]. The commutative analogue is the famous polynomial identity testing (PIT) problem, for which designing a deterministic polynomial time algorithm remains a major open question in derandomization and complexity theory. We remark that the corresponding algebraic variety $\mathrm{Sing}_{n,m}$, consisting of m-tuples in $\mathrm{Mat}(n)$ which span only singular matrices, recently has been shown to be not a null cone [65].

▶ **Example 2.4** (Generalized Kronecker quivers). The action of $G = \mathrm{GL}(n) \times \mathrm{GL}(n)$ on k-tuples of matrices $(X_1, \ldots, X_k)$ via $\pi(g, h)(X_1, \ldots, X_k) := (gX_1h^{-1}, \ldots, gX_kh^{-1})$ is sometimes also referred to as the left-right action. It can be obtained from action of Example 2.3 via the isomorphism $h \mapsto (h^{-1})^T$ of $\mathrm{GL}(n)$. This action is associated to the *generalized Kronecker quiver.*

▶ **Example 2.5** (Simultaneous conjugation). The action of the group $G = \mathrm{GL}(n)$ on k-tuples of matrices in $V = (\mathrm{Mat}(n))^d$ by $\pi(g)(X_1, \ldots, X_k) := (gX_1g^{-1}, \ldots, gX_kg^{-1})$ is associated to the quiver with a single vertex and k self-loops, briefly called *k-loop quiver.*

3 Computational problems and state of the art

In this section, we describe the main computational questions that are of interest for the optimization problems discussed in the previous section and then discuss what is known about them in the commutative and non-commutative worlds.

▶ **Problem 3.1** (Null cone membership). *Given (π, v), determine if v is in the null cone, i.e., if $\mathrm{cap}(v) = 0$. Equivalently, test if $0 \notin \Delta(v)$.*

The null cone membership problem for $\mathrm{GL}(n)$ is interesting only when the action $\pi(g)$ is given by rational functions in the $g_{i,j}$ rather than polynomials. This is completely analogous to the commutative case (e.g., the convex hull of weights ω_j with positive entries never contains the origin). In the important case that π is homogeneous, the null cone membership problem is interesting precisely when the total degree is zero, so that scalar multiples of the identity matrix act trivially. Thus, in this case the null cone membership problem for $G = \mathrm{GL}(n)$ is equivalent to the one for $G = \mathrm{SL}(n)$.

▶ **Problem 3.2** (Scaling). *Given (π, v, ε) such that $0 \in \Delta(v)$, output a group element $g \in G$ such that $\|\mathrm{spec}(\mu(g)v)\|_2 = \|\mu(\pi(g)v)\|_F \leq \varepsilon$.*

In particular, the following promise problem can be reduced to Problem 3.2: Given (π, v, ε), decide whether $0 \notin \Delta(v)$ under the promise that either $0 \in \Delta(v)$ or 0 is ε-far from $\Delta(v)$. In fact, there always exists $\varepsilon > 0$, depending only on the group action, such that this promise is satisfied! Thus, the null cone membership problem can always be reduced to the scaling problem (see Corollary 4.5 below).

One can develop a non-commutative duality theory [15, Section 3.4] showing that an efficient agorithm to minimize the norm on an orbit closure of a vector v (i.e., approximate the capacity of v) under the promise that $0 \in \Delta(v)$ results in an efficient algorithm for the scaling problem and hence for the null cone membership problem. This motivates our next computational problem.

▶ **Problem 3.3** (Norm minimization). *Given (π, v, ε) such that $\mathrm{cap}(v) > 0$, output a group element $g \in G$ such that $\log\|\pi(g)v\| - \log \mathrm{cap}(v) \leq \varepsilon$.*

We also consider the moment polytope membership problem for an arbitrary point $p \in \mathbb{Q}^n$.

▶ **Problem 3.4** (Moment polytope membership)**.** *Given (π, v, p), determine if $p \in \Delta(v)$.*

The moment polytope membership problem is more general than the null cone membership problem, but there is a reduction from the former to the latter via the "shifting trick" from [70, 11], which forms the basis of the algorithms for the moment polytope membership problem in [15]. As in the case of the null cone, we can consider a scaling version of the moment polytope membership problem.

▶ **Problem 3.5** (p-scaling)**.** *Given (π, v, p, ε) such that $p \in \Delta(v)$, output an element $g \in G$ such that $\|\mathrm{spec}(\mu(\pi(g)v)) - p\|_2 \leq \varepsilon$.*

The above problem has been referred to as *nonuniform scaling* [13] or, for operators, matrices and tensors, as *scaling with specified or prescribed marginals* [34]. The following problem can be reduced to Problem 3.5: Given (π, v, p, ε), decide whether $p \in \Delta(v)$ under the promise that either $p \in \Delta(v)$ or p is ε-far from $\Delta(v)$. One can combine the shifting trick with the non-commutive duality theory to show that there is a value $\varepsilon > 0$ with bitsize polynomial in the input size such that this is promise is always satisfied [15]. Thus, the moment polytope membership problem can be reduced to p-scaling.

There are several interesting input models for these problems. One could explicitly describe the weights $\omega_1, \ldots, \omega_m$ for an action of $\mathrm{T}(n)$ (Equation (2.1)) and then describe v as $\sum_{j=1}^m c_j v_j$ by describing the c_j's. The analogous description in the non-commutative world would be to describe the irreducible representations occuring in V. Alternately, one could give black box access to the function $\|\pi(g)v\|$, or to the moment map $\mu(\pi(g)v)$, etc. Sometimes π can be a non-uniform input as well, such as a fixed family of representations like the simultaneous left-right action Example 2.3 as done in [37]. The inputs p and ε will be given in their binary descriptions but we will see that some of the algorithms run in time polynomial in their unary descriptions.

▶ Remark 3.6 (Running time in terms of ε). By standard considerations about the bit complexity of the facets of the moment polytope, it can be shown that polynomial time algorithms for the scaling problems (Problems 3.2 and 3.5) result in polynomial time algorithms for the exact versions (Problems 3.1 and 3.4, respectively). Polynomial time requires, in particular, $\mathrm{poly}(\log(1/\varepsilon))$ dependence on ε; a $\mathrm{poly}(1/\varepsilon)$ dependence is only known to suffice in special cases.

3.1 Commutative groups and geometric programming

In the commutative case, the preceding problems are reformulations of well-studied optimization problems and much is known about them computationally. To see this, consider the action of $\mathrm{T}(n)$ as in Equation (2.1), and a vector $v = \sum_{j=1}^m c_j v_j$. It follows from Section 2.4 that v is in the null cone iff $0 \notin \Delta(v) = \mathrm{conv}\{\omega_j : c_j \neq 0\}$. Recall from Equation (2.2), the formulation for $\mathrm{cap}(v)$. Since this formulation is convex, it follows that, given $\omega_1, \ldots, \omega_m \in \mathbb{Z}^n$ (recall this is the description of π) and $c_1, \ldots, c_m \in \mathbb{Q}[i]$ (each entry described in binary), there is a polynomial-time algorithm for the null cone membership problem via linear programming [54, 52]. The same is true for the moment polytope membership problem. The capacity optimization problem is an instance of *(unconstrained) geometric programming*. The recent paper [18] describes interior-point methods for this, which run in polynomial time. Before [18], it was hard to find an exact reference for the existence of a polynomial time algorithm for geometric programming; however, is was known that polynomial time can be achieved using the ellipsoid algorithm as done for the same problem in slightly different settings in the papers [43, 74, 75]. There has been work in the oracle setting as well, in which one has oracle access to the function $\|\pi(g)v\|$. The advantage of the oracle setting is that one can handle exponentially large representations of $\mathrm{T}(n)$ when it is not possible to describe all the weights explicitly. A very general result of this form is proved in [75]. While not explicitly mentioned in [75], their techniques can also be used to design polynomial time algorithms for *commutative* null cone and moment polytope membership in the oracle setting. Thus, in the commutative case, Problems 3.1, 3.3, and 3.4 are well-understood.

3.2 Non-commutative actions

Comparatively very little is known in the non-commutative case. In the special case, where the group is fixed, polynomial time algorithms were given by the use of quantifier elimination (which is inefficient) and, more recently, by Mulmuley in [67, Theorem 8.5] through a purely algebraic approach. For instance, this applies to the settings of $V = \mathrm{Sym}^d \mathbb{C}^n$ or $V = \Lambda^d \mathbb{C}^n$ with the natural action by $\mathrm{SL}(n)$, where n is fixed.

For nonfixed groups, the only two non-trivial group actions for which there are known polynomial-time algorithms for null cone membership (Problem 3.1) are the *simultaneous conjugation* (Example 2.5) and the *left-right* action (Example 2.4). Approximate algorithms for null cone membership have been designed for the *tensor action* of products of $\mathrm{SL}(n)$'s [16]. However the running time is exponential in the binary description of ε (i.e., polynomial in $1/\varepsilon$). This is the reason the algorithm does not lead to a polynomial time algorithm for the exact null cone membership problem for the tensor action.

Moment polytope membership is already interesting for the polytope Δ in (2.4), the moment polytope of the entire representation V (not restricted to any orbit closure). Even here, efficient algorithms are only known in very special cases, such as for the Horn polytope (Example 2.1) [64, 68, 17]. The structural results in [8, 72, 78] characterize Δ in terms of linear inequalities (it is known that in general there are exponentially many). Mathematically, this is related to the asymptotic vanishing of certain representation-theoretic multiplicities [11, 20, 6] whose non-vanishing is in general NP-hard to decide [48]. In [12] it was proved that the membership problem for Δ is in NP $\cap$ coNP. As Δ and $\Delta(v)$ coincide for generic $v \in V$, this problem captures the moment polytope membership problem (Problem 3.4) for almost all vectors (all except those in a set of measure zero).

The study of Problem 3.4 in the noncommutative case focused on *Brascamp-Lieb polytopes* (which are affine slices of moment polytopes). The paper [38] solved the moment polytope membership problem in time depending polynomially on the *unary* complexity of the target point. In [13], efficient algorithms were designed for the p-scaling problem (Problem 3.5) for tensor actions, extending the earlier work of [34] for the simultaneous left-right action. The running times of both algorithms are $\mathrm{poly}(1/\varepsilon)$; for this reason both algorithms result in moment polytope algorithms depending exponentially on the binary bitsize of p, as in [38].

Regarding the approximate computation of the capacity (Problem 3.3), efficient algorithms were previously known only for the simultaneous left-right action. The paper [37] gave an algorithm to approximate the capacity in time polynomial in all of the input description except ε, on which it had dependence $\mathrm{poly}(1/\varepsilon)$. The paper [2] gave an algorithm that depended polynomially on the input description; it has running time dependence $\mathrm{poly}(\log(1/\varepsilon))$ on the error parameter ε.

In terms of algorithmic techniques, all prior works that were based on optimization methods fall into two categories. One is that of *alternating minimization* (which can be thought of as a large-step coordinate gradient descent, i.e., roughly speaking as a first order method). However, alternating minimization is limited in applicability to "multilinear" actions of products of $\mathrm{T}(n)$'s or $\mathrm{GL}(n)$'s, where the action is linear in each component so that it is easy to optimize over one component when fixing all the others. This is true for all the actions described above and hence explains the applicability of alternating minimization (in fact, in all the above examples, one can even get a closed-form expression for the group element that has to be applied in each alternating step). The second category are geodesic analogues of *box-constrained Newton's methods* (second order). Recently, [2] designed an algorithm tailored towards the specific case of the simultaneous left-right action (Example 2.3), but no second order algorithms were known for other group actions. However, many group actions of interest – from classical problems in invariant theory about symmetric forms to the important variant of Problem 2 in the introduction for fermions – are not multilinear nor can otherwise be captured by the left-right action, and no efficient algorithms were known. All this motivates the development of new techniques.

The paper [15] shows how these limitations can be overcome. Specifically, it provides both first and second order algorithms (geodesic variants of gradient descent and box-constrained Newton's method) that apply in great generality and identify the main structural parameters that control the running time of these algorithm. We now describe these contributions in more detail.

4 Algorithmic and structural results

4.1 Essential parameters and structural results

We define here the essential parameters related to the group action which, in addition to dictating the running times of our first and second order methods, control the relationships between the null cone, the norm of the moment map, and the capacity, i.e., between Problems 3.1–3.3. For details we refer to [15]

We saw in Section 2 that for all actions of $\mathrm{T}(n)$ on a vector space V, one can find a basis of V consisting of simultaneous eigenvectors of the matrices $\pi(g)$, $g \in \mathrm{T}(n)$. While this is in general impossible for non-commutative groups, one can still decompose V into building blocks known as irreducible subspaces (or subrepresentations).

For $\mathrm{GL}(n)$, these are uniquely characterized by nonincreasing sequences $\lambda \in \mathbb{Z}^n$; such sequences λ are in bijection with irreducible representations $\pi_\lambda \colon \mathrm{GL}(n) \to \mathrm{GL}(V_\lambda)$. We say that λ *occurs in* π if one of its irreducible subspaces is of type λ. If all the λ occuring in π have nonnegative entries, then the entries of the matrix $\pi(g)$ are polynomials in the entries of g. Such representations π are called *polynomial*, and if all λ occuring in π have sum exactly (resp. at most) d, then π is said to be a *homogeneous polynomial representation of degree (resp. at most)* d.

Now we can define the complexity measure which captures the smoothness of the optimization problems of interest. One can think of the following measure as a *norm* of the Lie algebra representation Π, hence the name *weight norm*.

▶ **Definition 4.1** (Complexity measure I: weight norm). *We define the* weight norm $N(\pi)$ *of an action* π *of* $\mathrm{GL}(n)$ *by* $N(\pi) := \max\{\|\lambda\|_2 : \lambda$ *occurs in* $\pi\}$*, where* $\|\cdot\|_2$ *denotes the Euclidean norm.*

Another use of the weight norm is to provide a bounding ball for the moment polytope: one can show that the moment polytope is contained in a Euclidean ball of radius $N(\pi)$. The weight norm is in turn controlled by the degree of a polynomial representation. More specifically, if π is a polynomial representation of $\mathrm{GL}(n)$ of degree at most d, then $N(\pi) \leq d$.

We now describe our second measure of complexity which will govern the running time bound for our second order algorithm. This parameter also features in Theorem 4.3 concerning quantitative non-commutative duality.

▶ **Definition 4.2** (Complexity measure II: weight margin). *The* weight margin $\gamma(\pi)$ *of an action* π *of* $\mathrm{GL}(n)$ *is the minimum Euclidean distance between the origin and the convex hull of any subset of the weights of* π *that does not contain the origin.*

Our running time bound will depend inversely on the weight margin. Two interesting examples with large (inverse polynomial) weight margin are the left-right action (Example 2.3) and simultaneous conjugation. The existing second order algorithm for the left-right action relied on the large weight margin of the action [2]. It is interesting that the simultaneous conjugation action (Example 2.5), the sole other interesting example of an action of a non-commutative group for which there are efficient algorithms for the null cone membership problem [71, 33, 28] (which have nothing to do with the weight margin), also happens to have large weight margin! On the other hand, the only generally applicable lower bound on the weight margin is $N(\pi)^{1-n}n^{-1}$, and indeed this exponential behavior is seen for the somewhat intractable 3-tensor action (Example 2.2), which has weight margin at most $2^{-n/3}$ and weight norm $\sqrt{3}$ [61, 35]. We arrange in a tabular form the above information about the weight margin and weight norm for various paradigmatic group actions in Table 1 (using a definition of the weight margin and weight norm that naturally generalizes the one given above for $\mathrm{GL}(n)$).

Table 1 Weight margin and norm for various representations.

Group action	**Weight margin $\gamma(\pi)$**	**Weight norm $N(\pi)$**
Matrix scaling	$\geq n^{-3/2}$; [63]	$\sqrt{2}$
Simultan. left-right action (Example 2.3)	$\geq n^{-3/2}$; [42]	$\sqrt{2}$
Quivers	$\geq (\sum_x n(x))^{-3/2}$	$\sqrt{2}$
Simultaneous conjugation (Example 2.5)	$\geq n^{-3/2}$	$\sqrt{2}$
3-tensor action (Example 2.2)	$\leq 2^{-n/3}$; [61, 35]	$\sqrt{3}$
Polynomial $\mathrm{GL}(n)$-action of degree d	$\geq d^{-n}dn^{-1}$	$\leq d$
Polynomial $\mathrm{SL}(n)$-action of degree d	$\geq (nd)^{-n}dn^{-1}$	$\leq d$

As the moment map is the gradient of the geodesically convex function $\log\|v\|$, it stands to reason that as $\mu(v)$ tends to zero, $\|v\|$ tends to the capacity $\mathrm{cap}(v)$. However, in order to use this relationship to obtain efficient algorithms, we need this to hold in a precise quantitative sense. To this end, we show in [15] the following fundamental relation between the capacity and the norm of the moment map, which is a quantitative strengthening of the Kempf-Ness result [53].

▶ **Theorem 4.3** (Noncommutative duality). *For* $v \in V \setminus \{0\}$ *we have*

$$1 - \frac{\|\mu(v)\|_F}{\gamma(\pi)} \leq \frac{\mathrm{cap}(v)^2}{\|v\|^2} \leq 1 - \frac{\|\mu(v)\|_F^2}{4N(\pi)^2}.$$

Equipped with these inequalities, it is easy to relate Problems 3.2 and 3.3.

▶ **Corollary 4.4.** *An output* g *for the norm minimization problem on input* (π, v, ε) *is a valid output for the scaling problem on input* $(\pi, v, N(\pi)\sqrt{8\varepsilon})$. *If* $\varepsilon/\gamma(\pi) < \frac{1}{2}$ *then an output* g *for the scaling problem on input* (π, v, ε) *is a valid output for the norm minimization problem on input* $(\pi, v, \frac{2\log(2)\varepsilon}{\gamma(\pi)})$.

Because $0 \in \Delta(v)$ if and only if $\mathrm{cap}(v) > 0$, Theorem 4.3 and Corollary 4.4 immediately yield the accuracy to which we must solve the scaling problem or norm minimization problem to solve the null cone membership problem:

▶ **Corollary 4.5.** *It holds that $0 \in \Delta(v)$ if and only if $\Delta(v)$ contains a point of norm smaller than $\gamma(\pi)$. In particular, solving the scaling problem with input $(\pi, v, \gamma(\pi)/2)$ or the norm minimization problem with $(\pi, v, \frac{1}{8}(\gamma(\pi)/2N(\pi))^2)$ suffices to solve the null cone membership problem for (π, v).*

In [15] we also provide analogues of the above corollaries for the moment polytope membership problem.

4.2 First order methods: structural results and algorithms

As discussed above, in order to approximately compute the capacity in the commutative case, one can just run a Euclidean gradient descent on the convex formulation in Equation (2.2). We will see that the gradient descent method naturally generalizes to the non-commutative setting. It is worth mentioning that there are several excellent sources of the analysis of gradient descent algorithms for geodesically convex functions (in the general setting of Riemannian manifolds and not just the group setting that we are interested in); see e.g., [77, 1, 83, 82, 73, 84] and references therein. The contribution in [15] is mostly in understanding the geometric properties (such as smoothness) of the optimization problems that we are concerned with, which allow us to carry out the classical analysis of Euclidean gradient descent in our setting and to obtain quantitative convergence rates, which are not present in previous work.

The natural analogue of gradient descent for the optimization problem $\mathrm{cap}(v)$ is the following: start with $g_0 = I$ and repeat, for T iterations and a suitable step size η:

$$g_{t+1} = e^{-\eta\mu(\pi(g_t)v)} g_t. \tag{4.1}$$

Finally, return the group element g among $g_0, \ldots, g_{T-1}$, which minimizes $\|\mu(\pi(g)v)\|_F$. A natural geometric parameter which governs the complexity (number of iterations T, step size η) of gradient descent is the *smoothness* of the function to be optimized. The smoothness parameter for actions of $\mathrm{T}(n)$ in Equation (2.1) can be shown to be $O(\max_{j\in[m]}\|\omega_j\|_2^2)$ (see, e.g., [75]), which is the square of the weight norm defined in Definition 4.1 for this action. We prove in [15] that, in general, the function $\log\|\pi(g)v\|$ is geodesically smooth, with a smoothness parameter which, analogously to the commutative case, is on the order of the square of the weight norm. We now state the running time for our geodesic gradient descent algorithm for Problem 3.2.

▶ **Theorem 4.6** (First order algorithm for scaling). *Fix a representation $\pi : \mathrm{GL}(n) \to \mathrm{GL}(V)$ and a unit vector $v \in V$ such that $\mathrm{cap}(v) > 0$ (i.e., v is not in the null cone). Then the above analogue (4.1) of gradient descent, with a number of iterations at most*

$$T = O\left(\frac{N(\pi)^2}{\varepsilon^2}\big|\log \mathrm{cap}(v)\big|\right),$$

outputs a group element $g \in \mathrm{GL}(n)$ satisfying $\|\mu(\pi(g)v)\|_F \le \varepsilon$.

The analysis of Theorem 4.6 relies on the smoothness of the function $F_v(g) := \log\|\pi(g)v\|$, which implies that

$$F_v(e^H g) \le F_v(g) + \mathrm{tr}\big[\mu\big(\pi(g)v\big)H\big] + N(\pi)^2\|H\|_F^2,$$

for all $g \in \mathrm{GL}(n)$ and for all Hermitian $H \in \mathrm{Herm}(n)$.

The paper [15] also describes and analyzes a first order algorithm for the p-scaling problem via the shifting trick.

4.3 Second order methods: structural results and algorithms

As mentioned in Section 3, the paper [2] (following the algorithms developed in [3, 23] for the commutative Euclidean case) developed a second order polynomial-time algorithm for approximating the capacity for the simultaneous left-right action (Example 2.3) with running time polynomial in the bit description of the approximation parameter ε. In [15] this algorithm is generalized to arbitrary groups and actions. It repeatedly optimizes quadratic Taylor expansions of the objective in a small neighbourhood. Such algorithms also go by the name "trust-region methods" in the Euclidean optimization literature [24]. The running time of this algorithm depends inversely on the weight margin defined in Definition 4.2.

▶ **Theorem 4.7** (Second-order algorithm for norm minimization)**.** *Fix a representation* $\pi : \mathrm{GL}(n) \to \mathrm{GL}(V)$ *and a unit vector* $v \in V$ *such that* $\mathrm{cap}(v) > 0$*. Put* $C := |\log \mathrm{cap}(v)|$*,* $\gamma := \gamma(\pi)$ *and* $N := N(\pi)$*. Then the second order algorithm in [15], for a suitably regularized objective function, outputs* $g \in G$ *satisfying* $\log \|\pi(g)v\| \leq \log \mathrm{cap}(v) + \varepsilon$ *with a number of iterations at most*

$$T = O\left(\frac{N\sqrt{n}}{\gamma}\left(C + \log\frac{n}{\varepsilon}\right)\log\frac{C}{\varepsilon}\right).$$

The two main structural parameters which govern the runtime of the second order algorithm are the *robustness* (controlled by the weight norm) and a *diameter bound* (controlled by the weight margin). The robustness of a function bounds third derivatives in terms of second derivatives, similarly to the well-known notion of self concordance (however, in contrast to the latter, the robustness is not scale-invariant). As a consequence of the robustness, one shows that the function $F_v(g) = \log\|\pi(g)v\|$ is sandwiched between two quadratic expansions in a small neighbourhood:

$$F(g) + \partial_{t=0}F(e^{tH}g) + \frac{1}{2e}\partial^2_{t=0}F(e^{tH}g) \leq F(e^{H}g) \leq F(g) + \partial_{t=0}F(e^{tH}g) + \frac{e}{2}\partial^2_{t=0}F(e^{tH}g)$$

for every $g \in \mathrm{GL}(n)$ and $H \in \mathrm{Herm}(n)$ such that $\|H\|_F \leq 1/(4N(\pi))$.

Another ingredient in the analysis of the second order algorithm is to prove the existence of "well-conditioned" approximate minimizers, i.e., $g_\star \in G$, with small condition number satisfying $\log\|\pi(g_\star)v\| \leq \log\mathrm{cap}(v) + \varepsilon$. The bound on the condition numbers of approximate minimizers helps us ensure that the algorithm's trajectory always lies in a compact region with the use of appropriate regularizers. As in [2], this "diameter bound" is obtained by designing a suitable gradient flow and bounding the (continuous) time it takes for it to converge. A crucial ingredient of this analysis is Theorem 4.3 relating capacity and norm of the moment map.

This gradient flow approach, which can be traced back to works in symplectic geometry [56], is the only one we know for proving diameter bounds in the non-commutative case. In contrast, in the commutative case several different methods are available (see, e.g., [74, 75]). It is an important open problem to develop alternative methods for diameter bounds in the non-commutative case, which will also lead to improved running time bounds for the second order algorithm.

Finally, we note that [15] also contains results bounding the running time of the obtained algorithms, beyond the number of oracle calls, in terms of the bitsize needed to describe the given action π and the given vector v.

5 Conclusion

We believe that extending this theory will be fruitful both from a mathematical and computational point of view. The paper [15] points to the following intriguing open problems and suggests further research directions.

1. Is the null cone membership problem for general group actions in P? A natural intermediate goal is to prove that they are in NP ∩ coNP. The quantitative duality theory developed in this paper makes such a result plausible. The same question may be asked about the moment polytope membership problem for general group actions [12].
2. Can we find more general classes of problems or group actions where our algorithms run in polynomial time? In view of the complexity parameters we have identified, it is of particular interest to understand when the *weight margin* is only inverse polynomially rather than exponentially small.
3. Interestingly, when restricted to the commutative case discussed in Section 3, our algorithms' guarantees do not match those of cut methods in the spirit of the ellipsoid algorithm. Can we extend non-commutative/geodesic optimization to include cut methods as well as interior point methods? The foundations we lay in extending first and second order methods to the non-commutative case makes one optimistic that similar extensions are possible of other methods in standard convex optimization. The paper [18] explicitly designs and analyses a polynomial time interior-point method in the commutative setting.
4. Can geodesic optimization lead to new efficient algorithms in combinatorial optimization? We know that it captures algorithmic problems like bipartite matching (and more generally matroid intersection). How about perfect matching in general graphs – is the Edmonds polytope a moment polytope of a natural group action?
5. Can geodesic optimization lead to new efficient algorithms in algebraic complexity and derandomization? We know that the null cone membership problem captures polynomial identity testing (PIT) in non-commuting variables. The variety corresponding to classical PIT is however *not* a null cone [65]. Can our algorithms be extended beyond null cones to membership in more general classes of varieties?

References

1 Pierre-Antoine Absil, Robert E. Mahony, and Rodolphe Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2008. URL: `http://press.princeton.edu/titles/8586.html`.

2 Zeyuan Allen-Zhu, Ankit Garg, Yuanzhi Li, Rafael Mendes de Oliveira, and Avi Wigderson. Operator scaling via geodesically convex optimization, invariant theory and polynomial identity testing. In Ilias Diakonikolas, David Kempe, and Monika Henzinger, editors, *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018*, pages 172–181. ACM, 2018. `doi:10.1145/3188745.3188942`.

3 Zeyuan Allen-Zhu, Yuanzhi Li, Rafael Mendes de Oliveira, and Avi Wigderson. Much faster algorithms for matrix scaling. In Chris Umans, editor, *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017*, pages 890–901. IEEE Computer Society, 2017. `doi:10.1109/FOCS.2017.87`.

4 Carlos Améndola, Kathlén Kohn, Philipp Reichenbach, and Anna Seigal. Invariant theory and scaling algorithms for maximum likelihood estimation, 2020. `arXiv:2003.13662`.

5 Michael F Atiyah. Convexity and commuting Hamiltonians. *Bulletin of the London Mathematical Society*, 14(1):1–15, 1982. `doi:10.1112/blms/14.1.1`.

6 Velleda Baldoni, Michèle Vergne, and M. Walter. Computation of dilated Kronecker coefficients. *J. Symb. Comput.*, 84:113–146, 2018. `doi:10.1016/j.jsc.2017.03.005`.

7 Prakash Belkale and Shrawan Kumar. Eigenvalue problem and a new product in cohomology of flag varieties. *Inventiones mathematicae*, 166:185–228, 2006. doi:10.1007/s00222-006-0516-x.

8 Arkady Berenstein and Reyer Sjamaar. Coadjoint orbits, moment polytopes, and the Hilbert–Mumford criterion. *Journal of the American Mathematical Society*, 13(2):433–466, 2000. doi:10.1090/S0894-0347-00-00327-1.

9 Nicole Berline, Michèle Vergne, and Michael Walter. The Horn inequalities from a geometric point of view. *L'Enseignement Mathématique*, 63:403–470, 2017. arXiv:1611.06917.

10 Stephen P. Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2014. doi:10.1017/CBO9780511804441.

11 Michel Brion. Sur l'image de l'application moment. In *Séminaire d'algebre Paul Dubreil et Marie-Paule Malliavin*, volume 1296 of *Lecture Notes in Mathematics*, pages 177–192. Springer, 1987.

12 Peter Bürgisser, Matthias Christandl, Ketan D. Mulmuley, and Michael Walter. Membership in moment polytopes is in NP and coNP. *SIAM J. Comput.*, 46(3):972–991, 2017. doi:10.1137/15M1048859.

13 Peter Bürgisser, Cole Franks, Ankit Garg, Rafael Mendes de Oliveira, Michael Walter, and Avi Wigderson. Efficient algorithms for tensor scaling, quantum marginals and moment polytopes. *CoRR*, abs/1804.04739, 2018. arXiv:1804.04739.

14 Peter Bürgisser, Cole Franks, Ankit Garg, Rafael Mendes de Oliveira, Michael Walter, and Avi Wigderson. Towards a theory of non-commutative optimization: Geodesic 1st and 2nd order methods for moment maps and polytopes. In David Zuckerman, editor, *60th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2019, Baltimore, Maryland, USA, November 9-12, 2019*, pages 845–861. IEEE Computer Society, 2019. doi:10.1109/FOCS.2019.00055.

15 Peter Bürgisser, Cole Franks, Ankit Garg, Rafael Mendes de Oliveira, Michael Walter, and Avi Wigderson. Towards a theory of non-commutative optimization: geodesic first and second order methods for moment maps and polytopes. *CoRR*, abs/1910.12375, 2019. arXiv:1910.12375.

16 Peter Bürgisser, Ankit Garg, Rafael Mendes de Oliveira, Michael Walter, and Avi Wigderson. Alternating minimization, scaling algorithms, and the null-cone problem from invariant theory. In Anna R. Karlin, editor, *9th Innovations in Theoretical Computer Science Conference, ITCS 2018, January 11-14, 2018, Cambridge, MA, USA*, volume 94 of *LIPIcs*, pages 24:1–24:20. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2018. doi:10.4230/LIPIcs.ITCS.2018.24.

17 Peter Bürgisser and Christian Ikenmeyer. Deciding positivity of Littlewood-Richardson coefficients. *SIAM J. Discret. Math.*, 27(4):1639–1681, 2013. doi:10.1137/120892532.

18 Peter Bürgisser, Yinan Li, Harold Nieuwboer, and Michael Walter. Interior-point methods for unconstrained geometric programming and scaling problems, 2020. arXiv:2008.12110.

19 Matthias Christandl, Brent Doran, Stavros Kousidis, and Michael Walter. Eigenvalue distributions of reduced density matrices. *Communications in Mathematical Physics*, 332(1):1–52, 2014. doi:10.1007/s00220-014-2144-4.

20 Matthias Christandl, Brent Doran, and Michael Walter. Computing multiplicities of Lie group representations. In *53rd Annual IEEE Symposium on Foundations of Computer Science, FOCS 2012, New Brunswick, NJ, USA, October 20-23, 2012*, pages 639–648. IEEE Computer Society, 2012. doi:10.1109/FOCS.2012.43.

21 Matthias Christandl, Aram W Harrow, and Graeme Mitchison. Nonzero Kronecker coefficients and what they tell us about spectra. *Communications in Mathematical Physics*, 270(3):575–585, 2007. doi:10.1007/s00220-006-0157-3.

22 Matthias Christandl and Graeme Mitchison. The spectra of quantum states and the Kronecker coefficients of the symmetric group. *Communications in Mathematical Physics*, 261(3):789–797, 2006. doi:10.1007/s00220-005-1435-1.

23 Michael B. Cohen, Aleksander Madry, Dimitris Tsipras, and Adrian Vladu. Matrix scaling and balancing via box constrained newton's method and interior point methods. In Chris Umans, editor, *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017*, pages 902–913. IEEE Computer Society, 2017. `doi:10.1109/FOCS.2017.88`.

24 Andrew R. Conn, Nicholas I. M. Gould, and Philippe L. Toint. *Trust Region Methods*. MOS-SIAM Series on Optimization. SIAM, 2000. `doi:10.1137/1.9780898719857`.

25 Sumit Daftuar and Patrick Hayden. Quantum state transformations and the Schubert calculus. *Annals of Physics*, 315(1):80–122, 2005. `doi:10.1016/j.aop.2004.09.012`.

26 Harm Derksen and Gregor Kemper. *Computational invariant theory*. Springer, 2015.

27 Harm Derksen and Visu Makam. Polynomial degree bounds for matrix semi-invariants. *Advances in Mathematics*, 310:44–63, 2017. `doi:10.1016/j.aim.2017.01.018`.

28 Harm Derksen and Visu Makam. Algorithms for orbit closure separation for invariants and semi-invariants of matrices, 2018. `arXiv:1801.02043`.

29 Harm Derksen and Visu Makam. Maximum likelihood estimation for matrix normal models via quiver representations, 2020. `arXiv:2007.10206`.

30 Harm Derksen, Visu Makam, and Michael Walter. Maximum likelihood estimation for tensor normal models via Castling transforms, 2020. `arXiv:2011.03849`.

31 Harm Derksen and Jerzy Weyman. *An introduction to quiver representations*, volume 184. American Mathematical Society, 2017.

32 Mathias Drton, Satoshi Kuriki, and Peter Hoff. Existence and uniqueness of the Kronecker covariance MLE, 2020. `arXiv:2003.06024`.

33 Michael A. Forbes and Amir Shpilka. Explicit noether normalization for simultaneous conjugation via polynomial identity testing. *Electron. Colloquium Comput. Complex.*, 20:33, 2013. URL: `http://eccc.hpi-web.de/report/2013/033`.

34 Cole Franks. Operator scaling with specified marginals. In Ilias Diakonikolas, David Kempe, and Monika Henzinger, editors, *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018*, pages 190–203. ACM, 2018. `doi:10.1145/3188745.3188932`.

35 Cole Franks and Philipp Reichenbach. Barriers for recent methods in geodesic optimization. Preprint, 2020. `arXiv:2102.06652`.

36 William Fulton. Eigenvalues, invariant factors, highest weights, and Schubert calculus. *Bulletin of the American Mathematical Society*, 37(3):209–249, 2000. `arXiv:math/9908012`.

37 Ankit Garg, Leonid Gurvits, Rafael Mendes de Oliveira, and Avi Wigderson. A deterministic polynomial time algorithm for non-commutative rational identity testing. In Irit Dinur, editor, *IEEE 57th Annual Symposium on Foundations of Computer Science, FOCS 2016, 9-11 October 2016, Hyatt Regency, New Brunswick, New Jersey, USA*, pages 109–117. IEEE Computer Society, 2016. `doi:10.1109/FOCS.2016.95`.

38 Ankit Garg, Leonid Gurvits, Rafael Mendes de Oliveira, and Avi Wigderson. Algorithmic and optimization aspects of brascamp-lieb inequalities, via operator scaling. In Hamed Hatami, Pierre McKenzie, and Valerie King, editors, *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, June 19-23, 2017*, pages 397–409. ACM, 2017. `doi:10.1145/3055399.3055458`.

39 Ankit Garg, Leonid Gurvits, Rafael Mendes de Oliveira, and Avi Wigderson. Operator scaling: Theory and applications. *Found. Comput. Math.*, 20(2):223–290, 2020. `doi:10.1007/s10208-019-09417-z`.

40 Ankit Garg, Christian Ikenmeyer, Visu Makam, Rafael Mendes de Oliveira, Michael Walter, and Avi Wigderson. Search problems in algebraic complexity, GCT, and hardness of generators for invariant rings. In Shubhangi Saraf, editor, *35th Computational Complexity Conference, CCC 2020, July 28-31, 2020, Saarbrücken, Germany (Virtual Conference)*, volume 169 of *LIPIcs*, pages 12:1–12:17. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020. `doi:10.4230/LIPIcs.CCC.2020.12`.

41 V. Guillemin and S. Sternberg. Convexity properties of the moment mapping. *Inventiones mathematicae*, 67:491–513, 1982.
42 Leonid Gurvits. Classical complexity and quantum entanglement. *J. Comput. Syst. Sci.*, 69(3):448–484, 2004. doi:10.1016/j.jcss.2004.06.003.
43 Leonid Gurvits. Combinatorial and algorithmic aspects of hyperbolic polynomials. *Electron. Colloquium Comput. Complex.*, (070), 2004. URL: http://eccc.hpi-web.de/eccc-reports/2004/TR04-070/index.html.
44 Leonid Gurvits. Hyperbolic polynomials approach to van der waerden/schrijver-valiant like conjectures: sharper bounds, simpler proofs and algorithmic applications. In Jon M. Kleinberg, editor, *Proceedings of the 38th Annual ACM Symposium on Theory of Computing, Seattle, WA, USA, May 21-23, 2006*, pages 417–426. ACM, 2006. doi:10.1145/1132516.1132578.
45 Leonid Gurvits and Peter N. Yianilos. The deflation-inflation method for certain semidefinite programming and maximum determinant completion problems. *Technical Report, NECI*, 1998.
46 Linus Hamilton and Ankur Moitra. The paulsen problem made simple. In Avrim Blum, editor, *10th Innovations in Theoretical Computer Science Conference, ITCS 2019, January 10-12, 2019, San Diego, California, USA*, volume 124 of *LIPIcs*, pages 41:1–41:6. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2019. doi:10.4230/LIPIcs.ITCS.2019.41.
47 David Hilbert. Über die vollen Invariantensysteme. *Math. Ann.*, 42:313–370, 1893.
48 Christian Ikenmeyer, Ketan D. Mulmuley, and Michael Walter. On vanishing of Kronecker coefficients. *Comput. Complex.*, 26(4):949–992, 2017. doi:10.1007/s00037-017-0158-y.
49 Gábor Ivanyos, Youming Qiao, and K. V. Subrahmanyam. Constructive non-commutative rank computation is in deterministic polynomial time. In Christos H. Papadimitriou, editor, *8th Innovations in Theoretical Computer Science Conference, ITCS 2017, January 9-11, 2017, Berkeley, CA, USA*, volume 67 of *LIPIcs*, pages 55:1–55:19. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2017. doi:10.4230/LIPIcs.ITCS.2017.55.
50 Gábor Ivanyos, Youming Qiao, and K. V. Subrahmanyam. Non-commutative Edmonds' problem and matrix semi-invariants. *Comput. Complex.*, 26(3):717–763, 2017. doi:10.1007/s00037-016-0143-x.
51 Valentine Kabanets and Russell Impagliazzo. Derandomizing polynomial identity tests means proving circuit lower bounds. *Comput. Complex.*, 13(1-2):1–46, 2004. doi:10.1007/s00037-004-0182-6.
52 Narendra Karmarkar. A new polynomial-time algorithm for linear programming. *Comb.*, 4(4):373–396, 1984. doi:10.1007/BF02579150.
53 George Kempf and Linda Ness. The length of vectors in representation spaces. In *Algebraic geometry*, pages 233–243. Springer, 1979.
54 Leonid G Khachiyan. A polynomial algorithm in linear programming. In *Doklady Academii Nauk SSSR*, volume 244, pages 1093–1096, 1979.
55 Frances Kirwan. Convexity properties of the moment mapping, III. *Inventiones mathematicae*, 77(3):547–552, 1984.
56 Frances Clare Kirwan. *Cohomology of quotients in symplectic and algebraic geometry*, volume 31. Princeton University Press, 1984.
57 Alexander Klyachko. Quantum marginal problem and representations of the symmetric group, 2004. arXiv:quant-ph/0409113.
58 Alexander A Klyachko. Stable bundles, representation theory and hermitian operators. *Selecta Mathematica, New Series*, 4(3):419–445, 1998. doi:10.1007/s000290050037.
59 A Knutson and T Tao. The honeycomb model of $\mathrm{GL}_n(\mathbb{C})$ tensor products I: Proof of the saturation conjecture. *Journal of the American Mathematical Society*, 12(4):1055–1090, 1999. arXiv:math/9807160.
60 B. Kostant. On convexity, the Weyl group and the Iwasawa decomposition. *Ann. scient. E.N.S*, 6:413–455, 1973.
61 V. M. Kravtsov. Combinatorial properties of noninteger vertices of a polytope in a three-index axial assignment problem. *Cybernetics and Systems Analysis*, 43(1):25–33, 2007.

62 Tsz Chiu Kwok, Lap Chi Lau, Yin Tat Lee, and Akshay Ramachandran. The Paulsen problem, continuous operator scaling, and smoothed analysis. In Ilias Diakonikolas, David Kempe, and Monika Henzinger, editors, *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018*, pages 182–189. ACM, 2018. doi:10.1145/3188745.3188794.

63 Nathan Linial, Alex Samorodnitsky, and Avi Wigderson. A deterministic strongly polynomial algorithm for matrix scaling and approximate permanents. In Jeffrey Scott Vitter, editor, *Proceedings of the Thirtieth Annual ACM Symposium on the Theory of Computing, Dallas, Texas, USA, May 23-26, 1998*, pages 644–652. ACM, 1998. doi:10.1145/276698.276880.

64 Jesús A. De Loera and Tyrrell B. McAllister. On the computation of Clebsch-Gordan coefficients and the dilation effect. *Exp. Math.*, 15(1):7–19, 2006. doi:10.1080/10586458.2006.10128948.

65 Visu Makam and Avi Wigderson. Singular tuples of matrices is not a null cone (and, the symmetries of algebraic varieties). *CoRR*, abs/1909.00857, 2019. arXiv:1909.00857.

66 Ketan Mulmuley. Geometric complexity theory V: equivalence between blackbox derandomization of polynomial identity testing and derandomization of noether's normalization lemma. In *53rd Annual IEEE Symposium on Foundations of Computer Science, FOCS 2012, New Brunswick, NJ, USA, October 20-23, 2012*, pages 629–638. IEEE Computer Society, 2012. doi:10.1109/FOCS.2012.15.

67 Ketan Mulmuley. Geometric complexity theory V: Efficient algorithms for Noether normalization. *Journal of the American Mathematical Society*, 30(1):225–309, 2017. arXiv:1209.5993.

68 Ketan D Mulmuley, Hariharan Narayanan, and Milind Sohoni. Geometric complexity theory III: on deciding nonvanishing of a Littlewood–Richardson coefficient. *Journal of Algebraic Combinatorics*, 36(1):103–110, 2012.

69 David Mumford. *Geometric invariant theory*. Springer-Verlag, 1965.

70 Linda Ness and David Mumford. A stratification of the null cone via the moment map. *American Journal of Mathematics*, 106(6):1281–1329, 1984. doi:10.2307/2374395.

71 Ran Raz and Amir Shpilka. Deterministic polynomial identity testing in non-commutative models. *Comput. Complex.*, 14(1):1–19, 2005. doi:10.1007/s00037-005-0188-8.

72 Nicolas Ressayre. Geometric invariant theory and the generalized eigenvalue problem. *Inventiones mathematicae*, 180(2):389–441, 2010. doi:10.1007/s00222-010-0233-3.

73 Hiroyuki Sato, Hiroyuki Kasai, and Bamdev Mishra. Riemannian stochastic variance reduced gradient algorithm with retraction and vector transport. *SIAM J. Optim.*, 29(2):1444–1472, 2019. doi:10.1137/17M1116787.

74 Mohit Singh and Nisheeth K. Vishnoi. Entropy, optimization and counting. In David B. Shmoys, editor, *Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 - June 03, 2014*, pages 50–59. ACM, 2014. doi:10.1145/2591796.2591803.

75 Damian Straszak and Nisheeth K. Vishnoi. Maximum entropy distributions: Bit complexity and stability. In Alina Beygelzimer and Daniel Hsu, editors, *Conference on Learning Theory, COLT 2019, 25-28 June 2019, Phoenix, AZ, USA*, volume 99 of *Proceedings of Machine Learning Research*, pages 2861–2891. PMLR, 2019. URL: http://proceedings.mlr.press/v99/straszak19a.html.

76 Bernd Sturmfels. *Algorithms in Invariant Theory*. Texts & Monographs in Symbolic Computation. Springer, 2008. doi:10.1007/978-3-211-77417-5.

77 Constantin Udriste. *Convex functions and optimization methods on Riemannian manifolds*, volume 297. Springer, 1994.

78 Michele Vergne and Michael Walter. Inequalities for moment cones of finite-dimensional representations. *Journal of Symplectic Geometry*, 15(4):1209–1250, 2017. doi:10.4310/JSG.2017.v15.n4.a8.

79 Frank Verstraete, Jeroen Dehaene, and Bart De Moor. Normal forms and entanglement measures for multipartite quantum states. *Physical Review A*, 68(1):012103, 2003. doi:10.1103/PhysRevA.68.012103.

80 Michael Walter. *Multipartite Quantum States and their Marginals*. PhD thesis, ETH Zurich, 2014. doi:10.3929/ethz-a-010250985.

81 Michael Walter, Brent Doran, David Gross, and Matthias Christandl. Entanglement polytopes: multiparticle entanglement from single-particle information. *Science*, 340(6137):1205–1208, 2013. doi:10.1126/science.1232957.

82 Hongyi Zhang, Sashank J. Reddi, and Suvrit Sra. Fast stochastic optimization on riemannian manifolds. *CoRR*, abs/1605.07147, 2016. arXiv:1605.07147.

83 Hongyi Zhang and Suvrit Sra. First-order methods for geodesically convex optimization. *CoRR*, abs/1602.06053, 2016. arXiv:1602.06053.

84 Hongyi Zhang and Suvrit Sra. Towards riemannian accelerated gradient methods. *CoRR*, abs/1806.02812, 2018. arXiv:1806.02812.