Invariants, polytopes, and optimization

Michael Walter

University of Amsterdam · QuSoft Research Center for Quantum Software · NWO

Lower Bounds in Computational Complexity Reunion
Berkeley, December 2019

based on joint work with Peter Bürgisser, Cole Franks, Ankit Garg, Rafael Oliveira, Avi Wigderson (ITCS'18, FOCS'18, FOCS'19)

Overview

There are algebraic and geometric problems in invariant theory that are amenable to numerical optimization algorithms over noncommut. groups.

Null cone & moment polytopes $\longleftrightarrow$ Norm minimization

These capture a wide range of surprising applications – from algebra and analysis to computer science and even quantum information.

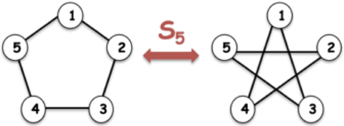
Plan for today:

1. Introduction to framework
2. Panorama of applications
3. Geodesic first-order algorithms

'Computational invariant theory ***without*** *computing invariants'*

Symmetries and group actions

Group actions mathematically model *symmetries* and *equivalence*.



Problem: How can we algorithmically and efficiently check equivalence?

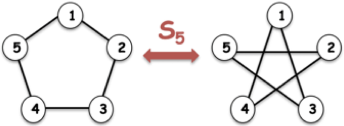
Interesting (and often difficult) problems with many applications:

- no polynomial-time algorithms are known for graph isomorphism.
- matrices equivalent under left-right action iff equal rank;
  but tensor rank is NP-hard.
- the 'flip' in geometric complexity theory: [Mulmuley]
  lower bounds from symmetry obstructions
- derandomizing PIT implies circuit lower bounds [Kabanets-Impagliazzo]

We will see many more examples in a moment. . .

Symmetries and group actions

Group actions mathematically model *symmetries* and *equivalence*.



Problem: How can we algorithmically and efficiently check equivalence?

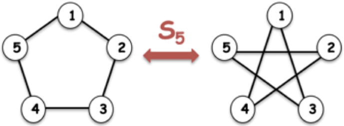
Interesting (and often difficult) problems with many applications:

- no polynomial-time algorithms are known for graph isomorphism.
- matrices equivalent under left-right action iff equal rank;
  but tensor rank is NP-hard.
- the 'flip' in geometric complexity theory: [Mulmuley]
  lower bounds from symmetry obstructions
- derandomizing PIT implies circuit lower bounds [Kabanets-Impagliazzo]

We will see many more examples in a moment. . .

Symmetries and group actions

Group actions mathematically model *symmetries* and *equivalence*.



Problem: How can we algorithmically and efficiently check equivalence?

Interesting (and often difficult) problems with many applications:

- no polynomial-time algorithms are known for graph isomorphism.
- matrices equivalent under left-right action iff equal rank; but tensor rank is NP-hard.
- the 'flip' in geometric complexity theory: lower bounds from symmetry obstructions [Mulmuley]
- derandomizing PIT implies circuit lower bounds [Kabanets-Impagliazzo]

We will see many more examples in a moment...

Setup and orbit problems

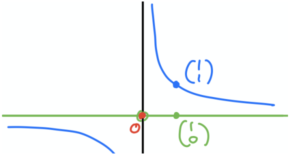
group $G \subseteq \mathrm{GL}_n(\mathbb{C})$, such as GL_n, SL_n, or $\mathrm{T}_n = \left(\ddots\right)$

action on $V = \mathbb{C}^m$ by linear transformations

orbits $Gv = \{g \cdot v \ : \ g \in G\}$ and their closures $\overline{Gv}$

Example: $G = \mathrm{GL}_1 = \mathbb{C}^*$, $V = \mathbb{C}^2$

$$g \cdot \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} gx \\ g^{-1}y \end{pmatrix}$$



Orbit equality problem: Given v_1 and v_2, is $Gv_1 = Gv_2$? *Robust version:*

Orbit closure intersection problem: Given v_1 and v_2, is $\overline{Gv_1} \cap \overline{Gv_2} \neq \emptyset$?

Null cone problem: Given v, is $0 \in \overline{Gv}$?

The last two can be solved via invariants (cf. Rafael's talk), but there are more efficient ways!

Setup and orbit problems

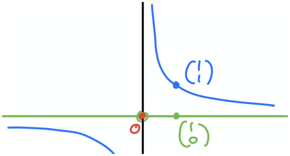
group $G \subseteq \mathrm{GL}_n(\mathbb{C})$, such as GL_n, SL_n, or $\mathrm{T}_n = \left(\begin{smallmatrix} \cdot & & \\ & \cdot & \\ & & \cdot \end{smallmatrix}\right)$

action on $V = \mathbb{C}^m$ by linear transformations

orbits $Gv = \{g \cdot v \ : \ g \in G\}$ and their closures $\overline{Gv}$

Example: $G = \mathrm{GL}_1 = \mathbb{C}^*$, $V = \mathbb{C}^2$

$$g \cdot \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} gx \\ g^{-1}y \end{pmatrix}$$



Orbit equality problem: Given v_1 and v_2, is $Gv_1 = Gv_2$? *Robust version:*

Orbit closure intersection problem: Given v_1 and v_2, is $\overline{Gv_1} \cap \overline{Gv_2} \neq \emptyset$?

Null cone problem: Given v, is $0 \in \overline{Gv}$?

The last two can be solved via invariants (cf. Rafael's talk), but there are more efficient ways!

Setup and orbit problems

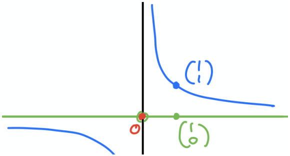
group $G \subseteq \mathrm{GL}_n(\mathbb{C})$, such as GL_n, SL_n, or $\mathrm{T}_n = \left(\ddots\right)$

action on $V = \mathbb{C}^m$ by linear transformations

orbits $Gv = \{g \cdot v \ : \ g \in G\}$ and their closures $\overline{Gv}$

Example: $G = \mathrm{GL}_1 = \mathbb{C}^*$, $V = \mathbb{C}^2$

$$g \cdot \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} gx \\ g^{-1}y \end{pmatrix}$$



Orbit equality problem: Given v_1 and v_2, is $Gv_1 = Gv_2$? *Robust version:*

Orbit closure intersection problem: Given v_1 and v_2, is $\overline{Gv_1} \cap \overline{Gv_2} \neq \emptyset$?

Null cone problem: Given v, is $0 \in \overline{Gv}$?

The last two can be solved via invariants (cf. Rafael's talk), but there are more efficient ways!

Example: Conjugation

$G = \mathrm{GL}_n, \quad V = \mathrm{Mat}_n, \quad g \cdot X = gXg^{-1}$

$$\begin{pmatrix} \lambda_1 & 1 & & \\ & \lambda_1 & 1 & \\ & & \lambda_1 & \\ & & & \ddots \end{pmatrix}$$

- X, Y are in *same orbit* iff same Jordan normal form
- X, Y have *intersecting orbit closures* iff same eigenvalues
- X is in *null cone* iff nilpotent

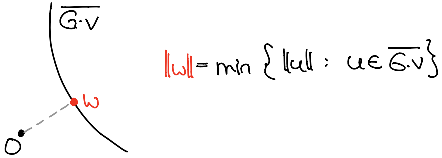
NB: The last two problems have a meaningful approximate version!

Null cone and norm minimization

We can translate the null cone problem into an optimization problem. Define capacity of v:

$$\mathrm{cap}(v) := \min_{u \in \overline{Gv}} \|u\| = \inf_{g \in G} \|g \cdot v\|$$

- clearly, $0 \in \overline{Gv}$ iff $\mathrm{cap}(v) = 0$



- generalizes Gurvits' notions of matrix, polynomial, operator capacity

Norm minimization problem: Given v, find $g \in G$ s. th. $\|g \cdot v\| \approx \mathrm{cap}(v)$.

Groups and derivatives

We want to minimize the function:

$$F_v \colon G \to \mathbb{R}, \quad F_v(g) := \log \|g \cdot v\|$$

First-order condition? How to define derivatives?

Consider $G = \mathrm{GL}_n$. Any invertible matrix g can be written as exponential:

$$\mathrm{GL}_n = \{g = e^A \;:\; A \in \mathrm{Mat}_n\}$$

Since $e^{At} = I + At + O(t^2)$, can think of A as a tangent direction:

Thus, $\partial_{t=0} F_v(e^{At})$ defines *derivative* at $g = I$ in direction A.

Similarly for general $G \subseteq \mathrm{GL}_n$ – only need to restrict allowed directions.

Groups and derivatives

We want to minimize the function:

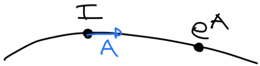
$$F_v \colon G \to \mathbb{R}, \quad F_v(g) := \log \|g \cdot v\|$$

First-order condition? How to define derivatives?

Consider $G = \mathrm{GL}_n$. Any invertible matrix g can be written as exponential:

$$\mathrm{GL}_n = \{g = e^A \ : \ A \in \mathrm{Mat}_n\}$$

Since $e^{At} = I + At + O(t^2)$, can think of A as a tangent direction:



Thus, $\partial_{t=0} F_v(e^{At})$ defines *derivative* at $g = I$ in direction A.

Similarly for general $G \subseteq \mathrm{GL}_n$ – only need to restrict allowed directions.

Groups and derivatives

We want to minimize the function:

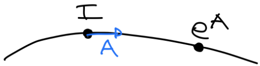
$$F_v\colon G \to \mathbb{R}, \quad F_v(g) := \log \|g \cdot v\|$$

First-order condition? How to define derivatives?

Consider $G = \mathrm{GL}_n$. Any invertible matrix g can be written as exponential:

$$\mathrm{GL}_n = \{g = e^A \; : \; A \in \mathrm{Mat}_n\}$$

Since $e^{At} = I + At + O(t^2)$, can think of A as a tangent direction:



Thus, $\partial_{t=0} F_v(e^{At})$ defines *derivative* at $g = I$ in direction A.

Similarly for general $G \subseteq \mathrm{GL}_n$ – only need to restrict allowed directions.

Norm minimization and its dual

We want to minimize the function:

$$F_v \colon G \to \mathbb{R}, \quad F_v(g) = \log \|g \cdot v\|$$

Its directional derivatives at $g = I$ are given by $\partial_{t=0} F_v(e^{At})$.

The corresponding *gradient* is known as the moment map:

$$\mu \colon V \setminus \{0\} \to \mathrm{Herm}_n, \quad \mathrm{tr}(\mu(v)A) = \partial_{t=0} F_v(e^{At}) \quad \forall A$$

- clearly, $\mu(g \cdot v) = 0$ if g is minimizer
- amazingly, also sufficient [Kempf-Ness]

Scaling problem: Given v, find $g \in G$ such that $\mu(g \cdot v) \approx 0$.

Norm minimization and its dual

We want to minimize the function:

$$F_v \colon G \to \mathbb{R}, \quad F_v(g) = \log \|g \cdot v\|$$

Its directional derivatives at $g = I$ are given by $\partial_{t=0} F_v(e^{At})$.

The corresponding *gradient* is known as the moment map:

$$\mu \colon V \setminus \{0\} \to \mathrm{Herm}_n, \quad \mathrm{tr}(\mu(v)A) = \partial_{t=0} F_v(e^{At}) \quad \forall A$$

- clearly, $\mu(g \cdot v) = 0$ if g is minimizer
- amazingly, also sufficient [Kempf-Ness]

Scaling problem: Given v, find $g \in G$ such that $\mu(g \cdot v) \approx 0$.

Norm minimization and its dual

We want to minimize the function:

$$F_v\colon G \to \mathbb{R}, \quad F_v(g) = \log \|g \cdot v\|$$

Its directional derivatives at $g = I$ are given by $\partial_{t=0} F_v(e^{At})$.

The corresponding *gradient* is known as the moment map:

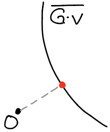
$$\mu\colon V \setminus \{0\} \to \mathrm{Herm}_n, \quad \mathrm{tr}(\mu(v)A) = \partial_{t=0} F_v(e^{At}) \quad \forall A$$

- clearly, $\mu(g \cdot v) = 0$ if g is minimizer
- amazingly, also sufficient [Kempf-Ness]

Scaling problem: Given v, find $g \in G$ such that $\mu(g \cdot v) \approx 0$.

Summary so far

$G \subseteq GL_n$ group, action on $V = \mathbb{C}^m$, $\mu\colon V \setminus \{0\} \to \mathrm{Herm}_n$ moment map



Null cone problem: Given v, is $0 \in \overline{Gv}$?

...and its relaxations:

Norm minimization problem: Given v, find $g \in G$ s. th. $\|g \cdot v\| \approx \mathrm{cap}(v)$.

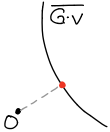
Scaling problem: Given $v \in V$, find $g \in G$ s. th. $\mu(g \cdot v) \approx 0$.

- The last two problems are dual to each other, and either can be used to solve null cone!
- But they also provide path to orbit closure intersection.

Useful model problems. Plausibly in P, and rich enough to have interesting applications. Let us look at some...

Summary so far

$G \subseteq \mathrm{GL}_n$ group, action on $V = \mathbb{C}^m$, $\mu\colon V \setminus \{0\} \to \mathrm{Herm}_n$ moment map



Null cone problem: Given v, is $0 \in \overline{Gv}$?

…and its relaxations:

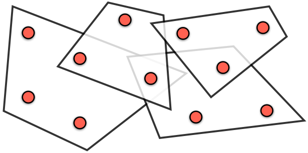
Norm minimization problem: Given v, find $g \in G$ s. th. $\|g \cdot v\| \approx \mathrm{cap}(v)$.

Scaling problem: Given $v \in V$, find $g \in G$ s. th. $\mu(g \cdot v) \approx 0$.

- The last two problems are dual to each other, and either can be used to solve null cone!
- But they also provide path to orbit closure intersection.

Useful model problems. Plausibly in P, and rich enough to have interesting applications. Let us look at some…

A panorama of applications



Example: Matrix scaling (raking, IPFP, ...)

Let X be matrix with nonnegative entries. A *scaling* of X is a matrix

$$Y = \begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_n \end{pmatrix} X \begin{pmatrix} b_1 & & \\ & \ddots & \\ & & b_n \end{pmatrix} \qquad (a_1, \dots, b_n > 0).$$

A matrix is called *doubly stochastic (d.s.)* if row & column sums are 1.

Matrix scaling: Given X, $\exists$ (approximately) d.s. scalings?

Permanent: ... iff $\mathrm{per}(X) > 0$!

- ... iff $\exists$ bipartite perfect matching in support of X
- can be decided in polynomial time
- find scalings by alternatingly fixing rows & columns ☺ [Sinkhorn]
- convergence controlled by permanent [Linial et al]

Connections to statistics, complexity, combinatorics, geometry, numerics, ...

Example: Matrix scaling (raking, IPFP, ...)

Let X be matrix with nonnegative entries. A *scaling* of X is a matrix

$$Y = \begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_n \end{pmatrix} X \begin{pmatrix} b_1 & & \\ & \ddots & \\ & & b_n \end{pmatrix} \qquad (a_1, \dots, b_n > 0).$$

A matrix is called *doubly stochastic (d.s.)* if row & column sums are 1.

> Matrix scaling: Given X, $\exists$ (approximately) d.s. scalings?

Permanent: ... iff $\operatorname{per}(X) > 0$!

- ... iff $\exists$ bipartite perfect matching in support of X
- can be decided in polynomial time
- find scalings by alternatingly fixing rows & columns ☺ [Sinkhorn]
- convergence controlled by permanent [Linial et al]

Connections to statistics, complexity, combinatorics, geometry, numerics, ...

Example: Matrix scaling (raking, IPFP, ...)

Let X be matrix with nonnegative entries. A *scaling* of X is a matrix

$$Y = \begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_n \end{pmatrix} X \begin{pmatrix} b_1 & & \\ & \ddots & \\ & & b_n \end{pmatrix} \qquad (a_1, \dots, b_n > 0).$$

A matrix is called *doubly stochastic (d.s.)* if row & column sums are 1.

> Matrix scaling: Given X, $\exists$ (approximately) d.s. scalings?

Pe

$$V = \mathrm{Mat}_n, \quad G = \mathrm{T}_n \times \mathrm{T}_n, \quad (g_1, g_2)v = g_1 v g_2.$$

$$\mu \colon V \setminus \{0\} \to \mathbb{R}^n \oplus \mathbb{R}^n$$

$$\mu(v) = (\text{row sums}, \text{column sums}) \text{ of } X_{i,j} = \frac{|v_{i,j}|^2}{\|v\|^2}$$

orn]

t al]

Connections to statistics, complexity, combinatorics, geometry, numerics, ...

Example: Schur-Horn theorem

Let $\lambda_1 \geqslant \cdots \geqslant \lambda_n$ and $\delta_1, \ldots, \delta_n$ be integers.

Given λ and δ, $\exists$ Hermitian matrix with spectrum λ and diagonal δ?

$$U \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} U^* = \begin{pmatrix} \delta_1 & \star & \star \\ \star & \ddots & \star \\ \star & \star & \delta_n \end{pmatrix}$$

Schur-Horn theorem: . . . iff δ in permutahedron generated by λ, i.e., in $\text{conv}(S_n \cdot \lambda)$!

[Nonenmacher, 2008]

Starting point for celebrated convexity results in symplectic geometry [Kostant, Atiyah, Guillemin-Sternberg, Duistermaat-Heckman, Mumford, Kirwan, . . .]

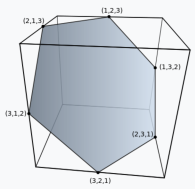
Example: Schur-Horn theorem

Let $\lambda_1 \geqslant \cdots \geqslant \lambda_n$ and $\delta_1, \ldots, \delta_n$ be integers.

Given λ and δ, $\exists$ Hermitian matrix with spectrum λ and diagonal δ?

$$U \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} U^* = \begin{pmatrix} \delta_1 & \star & \star \\ \star & \ddots & \star \\ \star & \star & \delta_n \end{pmatrix}$$

Schur-Horn theorem: ...iff δ in permutahedron generated by λ, i.e., in $\mathrm{conv}(S_n \cdot \lambda)$!



[Nonenmacher, 2008]

Starting point for celebrated convexity results in symplectic geometry [Kostant, Atiyah, Guillemin-Sternberg, Duistermaat-Heckman, Mumford, Kirwan, ...]

Example: Laurent polynomials

$\mathrm{T}_n = \left(\begin{smallmatrix} \cdot & \\ & \cdot \\ & & \cdot \end{smallmatrix} \right)$ acts on Laurent polynomials in n variables by scaling:

$$P = \sum_\omega p_\omega Z^\omega \qquad \Rightarrow \qquad g \cdot P = \sum_\omega p_\omega g^\omega Z^\omega$$

Capacity:

$$\mathrm{cap}(P)^2 = \inf_{g \in \mathrm{T}_n} \sum_\omega |p_\omega|^2 \, |g^\omega|^2 = \inf_{x \in \mathbb{R}^n} \sum_\omega |p_\omega|^2 \, e^{x \cdot \omega}$$

- geometric programming
- $\mathrm{cap}(P) = 0 \quad \text{iff} \quad 0 \notin \Delta(P) := \mathrm{conv}\,\{\omega : p_\omega \neq 0\}$

Moment map:

$$\mu(P) = \frac{\sum_\omega |p_\omega|^2 \, \omega}{\sum_\omega |p_\omega|^2}$$

- any point in $\Delta(P)$ can be obtained from scaling of P (approx.)

Example: Laurent polynomials

$\mathrm{T}_n = \begin{pmatrix} \cdot & & \\ & \cdot & \\ & & \cdot \end{pmatrix}$ acts on Laurent polynomials in n variables by scaling:

$$P = \sum_\omega p_\omega Z^\omega \qquad \Rightarrow \qquad g \cdot P = \sum_\omega p_\omega g^\omega Z^\omega$$

Capacity:

$$\mathrm{cap}(P)^2 = \inf_{g \in \mathrm{T}_n} \sum_\omega |p_\omega|^2 \, |g^\omega|^2 = \inf_{x \in \mathbb{R}^n} \sum_\omega |p_\omega|^2 \, e^{x \cdot \omega}$$

- geometric programming
- $\mathrm{cap}(P) = 0 \quad \text{iff} \quad 0 \notin \Delta(P) := \mathrm{conv}\,\{\omega : p_\omega \neq 0\}$

Moment map:

$$\mu(P) = \frac{\sum_\omega |p_\omega|^2 \, \omega}{\sum_\omega |p_\omega|^2}$$

- any point in $\Delta(P)$ can be obtained from scaling of P (approx.)

Example: Laurent polynomials

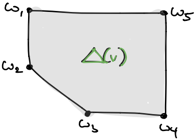
$\mathrm{T}_n = \begin{pmatrix} \cdot & \\ & \cdot \end{pmatrix}$ acts on Laurent polynomials in n variables by scaling:

$$P = \sum_\omega p_\omega Z^\omega \qquad \Rightarrow \qquad g \cdot P = \sum_\omega p_\omega g^\omega Z^\omega$$

Capacity:

$$\operatorname{cap}(P)^2 = \inf_{g \in \mathrm{T}_n} \sum_\omega |p_\omega|^2 \, |g^\omega|^2 = \inf_{x \in \mathbb{R}^n} \sum_\omega |p_\omega|^2 \, e^{x \cdot \omega}$$

- geometric programming
- $\operatorname{cap}(P) = 0$ iff $0 \notin \Delta(P) := \operatorname{conv}\{\omega : p_\omega \neq 0\}$



Newton polytope

Moment map:

$$\mu(P) = \frac{\sum_\omega |p_\omega|^2 \, \omega}{\sum_\omega |p_\omega|^2}$$

- any point in $\Delta(P)$ can be obtained from scaling of P (approx.)

Example: Laurent polynomials

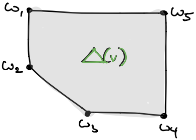
$\mathrm{T}_n = \left(\ddots\right)$ acts on Laurent polynomials in n variables by scaling:

$$P = \sum_\omega p_\omega Z^\omega \qquad \Rightarrow \qquad g \cdot P = \sum_\omega p_\omega g^\omega Z^\omega$$

Capacity:

$$\operatorname{cap}(P)^2 = \inf_{g \in \mathrm{T}_n} \sum_\omega |p_\omega|^2 \, |g^\omega|^2 = \inf_{x \in \mathbb{R}^n} \sum_\omega |p_\omega|^2 \, e^{x \cdot \omega}$$

- geometric programming
- $\operatorname{cap}(P) = 0 \quad \text{iff} \quad 0 \notin \Delta(P) := \operatorname{conv}\{\omega : p_\omega \neq 0\}$



Newton polytope

Moment map:

$$\mu(P) = \frac{\sum_\omega |p_\omega|^2 \, \omega}{\sum_\omega |p_\omega|^2}$$

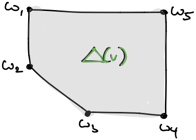
- any point in $\Delta(P)$ can be obtained from scaling of P (approx.)

Moment polytopes

- For $G = \mathrm{T}_n$, we saw on the previous slide that

$$\Delta(v) = \overline{\mu(Gv)} \subset \mathbb{R}^n$$

is a convex polytope.



- For $G = \mathrm{GL}_n$, get *magically* a convex polytope: [Mumford, Kirwan, ...]

$$\Delta(v) = \overline{\{\mathrm{spec}(\mu(g \cdot v)) : g \in G\}} \subset \mathbb{R}^n$$

These polytopes are known as moment polytopes.

Moment polytope problem: Given v and λ, is $\lambda \in \Delta(v)$?

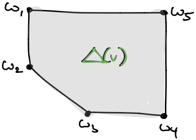
Even interesting when *not* restricted to orbit.

Moment polytopes

- For $G = \mathrm{T}_n$, we saw on the previous slide that

$$\Delta(v) = \overline{\mu(Gv)} \subset \mathbb{R}^n$$

is a convex polytope.



- For $G = \mathrm{GL}_n$, get *magically* a convex polytope: [Mumford, Kirwan, . . .]

$$\Delta(v) = \overline{\{\mathrm{spec}(\mu(g \cdot v)) : g \in G\}} \subset \mathbb{R}^n$$

These polytopes are known as moment polytopes.

Moment polytope problem: Given v and λ, is $\lambda \in \Delta(v)$?

Even interesting when *not* restricted to orbit.

Example: Horn problem

Let $\alpha_1 \geqslant \ldots \geqslant \alpha_n$, $\beta_1 \geqslant \ldots \geqslant \beta_n$, $\gamma_1 \geqslant \ldots \geqslant \gamma_n$ be integers.

Horn problem: When $\exists$ Hermitian $n \times n$ matrices A, B, C with spectrum α, β, γ such that $A + B = C$?

- exponentially many linear inequalities on α, β, γ [Horn]
- e.g., $\alpha_1 + \beta_1 \geqslant \gamma_1$

Knutson-Tao: ... iff *Littlewood-Richardson coefficient* $c^{\gamma}_{\alpha,\beta} > 0$

- counts multiplicities in representation theory, combinatorial gadgets, integer points in polytopes, ...
- poly-time algorithm [Mulmuley]
- can find A, B, C by natural algorithm [Franks]

Motivation for Mulmuley's positivity hypotheses.

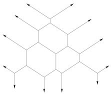
Example: Horn problem

Let $\alpha_1 \geqslant \ldots \geqslant \alpha_n$, $\beta_1 \geqslant \ldots \geqslant \beta_n$, $\gamma_1 \geqslant \ldots \geqslant \gamma_n$ be integers.

> Horn problem: When $\exists$ Hermitian $n \times n$ matrices A, B, C with spectrum α, β, γ such that $A + B = C$?

- exponentially many linear inequalities on α, β, γ [Horn]
- e.g., $\alpha_1 + \beta_1 \geqslant \gamma_1$

Knutson-Tao: …iff *Littlewood-Richardson coefficient* $c^{\gamma}_{\alpha,\beta} > 0$



- counts multiplicities in representation theory, combinatorial gadgets, integer points in polytopes, …
- poly-time algorithm [Mulmuley]
- can find A, B, C by natural algorithm [Franks]

Motivation for Mulmuley's positivity hypotheses.

Example: Left-right action and noncommutative PIT

Let $X = (X_1, \ldots, X_d)$ be a tuple of matrices. A *scaling* of X is a tuple

$$Y = (gX_1h^{-1}, \ldots, gX_dh^{-1}) \qquad (g, h \in \mathrm{GL}_n)$$

Say X is *quantum doubly stochastic* if $\sum_k X_k X_k^* = \sum_k X_k^* X_k = I$.

Operator scaling: Given X, $\exists$ (approx.) quantum d.s. scalings?

Non-commutative PIT: ...iff $\exists$ matrices Y_k s. th. $\sum_k Y_k \otimes X_k$ invertible.

- can solve in deterministic poly-time [Garg et al, cf. Ivanyos et al]
- when Y_k restricted to scalars: PIT for symbolic determinants ↯

Many further connections (Brascamp-Lieb inequalities, Paulsen problem, ...).

Example: Left-right action and noncommutative PIT

Let $X = (X_1, \ldots, X_d)$ be a tuple of matrices. A *scaling* of X is a tuple

$$Y = (gX_1h^{-1}, \ldots, gX_dh^{-1}) \qquad (g, h \in \mathrm{GL}_n)$$

Say X is *quantum doubly stochastic* if $\sum_k X_k X_k^* = \sum_k X_k^* X_k = I$.

> Operator scaling: Given X, $\exists$ (approx.) quantum d.s. scalings?

Non-commutative PIT: . . . iff $\exists$ matrices Y_k s. th. $\sum_k Y_k \otimes X_k$ invertible.

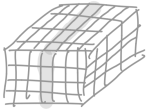
- can solve in deterministic poly-time [Garg et al, cf. Ivanyos et al]
- when Y_k restricted to scalars: PIT for symbolic determinants ↯

Many further connections (Brascamp-Lieb inequalities, Paulsen problem, . . .).

Example: Tensors and quantum marginals

Let $X \in \mathbb{C}^{n_1} \otimes \cdots \otimes \mathbb{C}^{n_d}$ be a tensor. A *scaling* of X is a tensor of the form

$$Y = (g_1 \otimes \ldots \otimes g_d)X \qquad (g_k \in \mathrm{GL}_{n_k})$$



Consider $\rho_k = X_k X_k^*$, where X_k is k-th flattening of X.

(In quantum mechanics, X describes joint state of d particles and ρ_k marginal of k-th particle.)

Tensor scaling problem: Given X, which $(\rho_1, \ldots, \rho_d)$ can be obtained by scaling?

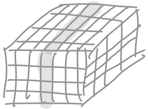
- eigenvalues form convex polytopes
- exponentially many vertices and faces
- characterized by asymptotics of *Kronecker coefficients*
  NP-hard to determine if nonzero [Ikenmeyer-Mulmuley-W]

Key challenge: Can we find efficient algorithmic description?

Example: Tensors and quantum marginals

Let $X \in \mathbb{C}^{n_1} \otimes \cdots \otimes \mathbb{C}^{n_d}$ be a tensor. A *scaling* of X is a tensor of the form

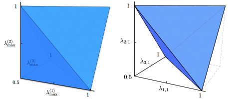
$$Y = (g_1 \otimes \ldots \otimes g_d)X \qquad (g_k \in \mathrm{GL}_{n_k})$$



Consider $\rho_k = X_k X_k^*$, where X_k is k-th flattening of X.

(In quantum mechanics, X describes joint state of d particles and ρ_k marginal of k-th particle.)

Tensor scaling problem: Given X, which $(\rho_1, \ldots, \rho_d)$ can be obtained by scaling?

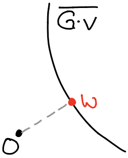


- eigenvalues form convex polytopes
- exponentially many vertices and faces
- characterized by asymptotics of *Kronecker coefficients*

  NP-hard to determine if nonzero [Ikenmeyer-Mulmuley-W]

Key challenge: Can we find efficient algorithmic description?

Geodesic first-order algorithms for norm minimization and scaling



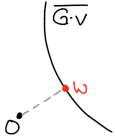
Non-commutative optimization duality

Recall $F_v(g) = \log \|g \cdot v\|$ and $\mu(v)$ is its gradient at $g = I$.

We discussed that the following *optimization problems* are equivalent:

$$\boxed{\inf_{g \in G} \; F_v(g)} \iff \boxed{\inf_{g \in G} \|\mu(g \cdot v)\|}$$

[Kempf-Ness]



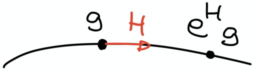
- primal: norm minimization, dual: scaling problem
- non-commutative version of linear programming duality

We developed quantitative duality theory and 1st & 2nd order methods.

Why does the duality hold at all? F_v is convex along geodesics!

Geodesic convexity and smoothness

For simplicity, $G = \mathrm{GL}_n$. Consider geodesics $\gamma(t) = e^{tH} g$ for $H \in \mathrm{Herm}_n$.



Proposition: F_v satisfies the following properties along these geodesics:

1. convexity: $\partial^2_{t=0} F_v(\gamma(t)) \geqslant 0$
2. smoothness: $\partial^2_{t=0} F_v(\gamma(t)) \leqslant 2N^2 \|H\|^2$

N is typically small, upper-bounded by degree of action.

Smoothness implies that

$$F_v(e^H g) \leqslant F_v(g) + \mathrm{tr}(\mu(v) H) + N^2 \|H\|^2.$$

Thus, gradient descent makes progress if steps not too large!

First-order algorithm: geodesic gradient descent

Given v, want to find $w = g \cdot v$ with $\|\mu(w)\| \leqslant \varepsilon$.

Algorithm: Start with $g = I$. For $t = 1, \ldots, T$:
Compute moment map $\mu(w)$ of $w = g \cdot v$. If norm ε-small, **stop**.
Otherwise, replace g by $e^{-\eta\mu(w)}g$. $\eta > 0$ *suitable step size*

Theorem

Let $v \in V$ be a vector with $\mathrm{cap}(v) > 0$. Then the algorithm outputs $g \in G$ such that $\|\mu(w)\| \leqslant \varepsilon$ within $T = \frac{4N^2}{\varepsilon^2} \log \frac{\|v\|}{\mathrm{cap}(v)}$ iterations.

- Algorithm runs in time $\mathrm{poly}(\frac{1}{\varepsilon}, \text{input size})$.
- Algorithm solves null cone problem if ε sufficiently small!

Peter Bürgisser will explain this in more detail.

First-order algorithm: geodesic gradient descent

Given v, want to find $w = g \cdot v$ with $\|\mu(w)\| \leqslant \varepsilon$.

Algorithm: Start with $g = I$. For $t = 1, \ldots, T$:
Compute moment map $\mu(w)$ of $w = g \cdot v$. If norm ε-small, **stop**.
Otherwise, replace g by $e^{-\eta\mu(w)} g$. $\eta > 0$ *suitable step size*

Theorem

Let $v \in V$ be a vector with $\operatorname{cap}(v) > 0$. Then the algorithm outputs $g \in G$ such that $\|\mu(w)\| \leqslant \varepsilon$ within $T = \frac{4N^2}{\varepsilon^2} \log \frac{\|v\|}{\operatorname{cap}(v)}$ iterations.

- Algorithm runs in time $\operatorname{poly}(\frac{1}{\varepsilon}, \text{input size})$.
- Algorithm solves null cone problem if ε sufficiently small!

Peter Bürgisser will explain this in more detail.

Analysis of algorithm

"Unless moment map ε-small, replace g by $e^{-\eta\mu(w)}g$."

To obtain rigorous algorithm, need to show *progress in each step*:

$$F_v(g_{\text{new}}) \leqslant F_v(g) - c$$

Then, $\log\|v\| - Tc \geqslant \log\operatorname{cap}(v)$ bounds the number of steps T.

Progress follows from smoothness:

$$F_v(e^H g) \leqslant F_v(g) + \operatorname{tr}(\mu(v)H) + N^2\|H\|^2$$

If we plug in $H = -\eta\mu(w)$ then

$$F_v(g_{\text{new}}) \leqslant F_v(g) - \eta\|\mu(w)\|^2 + N^2\eta^2\|\mu(w)\|^2.$$

Thus, if we choose $\eta = 1/2N^2$ then we obtain

$$F_v(g_{\text{new}}) \leqslant F_v(g) - \frac{1}{4N^2}\|\mu(w)\|^2 \leqslant F_v(g) - \frac{\varepsilon^2}{4N^2}. \qquad \square$$

Analysis of algorithm

"Unless moment map ε-small, replace g by $e^{-\eta\mu(w)}g$."

To obtain rigorous algorithm, need to show *progress in each step*:

$$F_v(g_{\text{new}}) \leqslant F_v(g) - c$$

Then, $\log\|v\| - Tc \geqslant \log\operatorname{cap}(v)$ bounds the number of steps T.

Progress follows from smoothness:

$$F_v(e^H g) \leqslant F_v(g) + \operatorname{tr}(\mu(v)H) + N^2\|H\|^2$$

If we plug in $H = -\eta\mu(w)$ then

$$F_v(g_{\text{new}}) \leqslant F_v(g) - \eta\|\mu(w)\|^2 + N^2\eta^2\|\mu(w)\|^2.$$

Thus, if we choose $\eta = 1/2N^2$ then we obtain

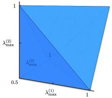
$$F_v(g_{\text{new}}) \leqslant F_v(g) - \frac{1}{4N^2}\|\mu(w)\|^2 \leqslant F_v(g) - \frac{\varepsilon^2}{4N^2}. \qquad \square$$

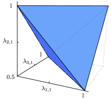
How about moment polytopes?

Recall:

Moment polytope problem: Given v and λ, is $\lambda \in \Delta(v)$?

- v in null cone $\Leftrightarrow 0 \notin \Delta(v)$
- how to reduce to $\lambda = 0$?





Shifting trick:

- Laurent polynomials: simply shift exponents $\omega \mapsto \omega - \lambda$
- If G noncommutative, more involved, need randomization [Mumford, Brion]

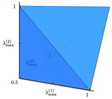
Result: *Randomized* first-order algorithm for moment polytopes.

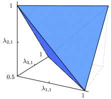
How about moment polytopes?

Recall:

Moment polytope problem: Given v and λ, is $\lambda \in \Delta(v)$?

- v in null cone $\Leftrightarrow 0 \notin \Delta(v)$
- how to reduce to $\lambda = 0$?



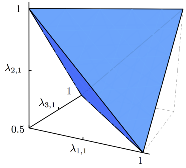


Shifting trick:

- Laurent polynomials: simply shift exponents $\omega \mapsto \omega - \lambda$
- If G noncommutative, more involved, need randomization [Mumford, Brion]

Result: Randomized first-order algorithm for moment polytopes.

Summary and outlook



Null cone & moment polytopes

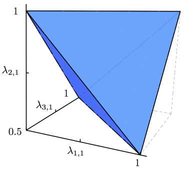
$\updownarrow$ duality

Norm minimization

Effective numerical algorithms for null cone and moment polytope problems, based on geodesic convex optimization and invariant theory, with a wide range of applications.

After the break, Peter Bürgisser will discuss the noncommutative duality theory in more detail and explain how to design second-order algorithms.

Summary and outlook



Null cone & moment polytopes

$\updownarrow$ duality

Norm minimization

Effective numerical algorithms for null cone and moment polytope problems, based on geodesic convex optimization and invariant theory, with a wide range of applications. *Many exciting directions:*

- Polynomial-time algorithms in all cases?
- Can we design geodesic interior point methods?
- Tensors in applications are often structured. Implications?
- What exponentially large polytopes can be efficiently captured?
- What are the tractable isomorphism problems? $\mathbb{C} \rightsquigarrow \mathbb{F}$?

Thank you for your attention!