Invariants, Algorithms, and Optimization

Michael Walter

University of Amsterdam NWO

CMI Webinar Series on Recent Progress in GCT
June 2020

based on joint works 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.

$$\boxed{\text{Null cone \& moment polytopes}} \quad \longleftrightarrow \quad \boxed{\text{Norm minimization}}$$

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

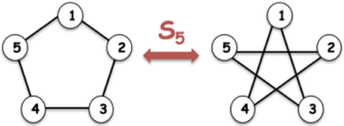
Plan for today:

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?

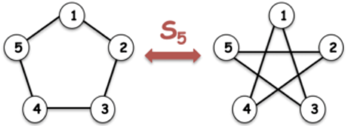
Interesting (and often difficult) problems with many applications:

- computing *normal forms*, describing *moduli spaces* and *invariants*...
- no polynomial-time algorithms are known for graph isomorphism.
- matrices equivalent under row and column operations iff equal rank; but tensor rank is NP-hard.
- 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?

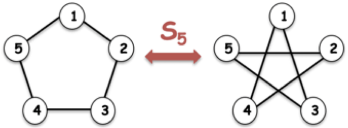
Interesting (and often difficult) problems with many applications:

- computing *normal forms*, describing *moduli spaces* and *invariants*...
- no polynomial-time algorithms are known for graph isomorphism.
- matrices equivalent under row and column operations iff equal rank; but tensor rank is NP-hard.
- 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?

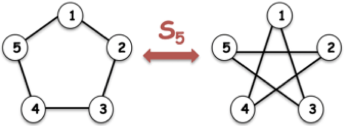
Interesting (and often difficult) problems with many applications:

- computing *normal forms*, describing *moduli spaces* and *invariants*...
- no polynomial-time algorithms are known for graph isomorphism.
- matrices equivalent under row and column operations iff equal rank; but tensor rank is NP-hard.
- 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:

- computing *normal forms*, describing *moduli spaces* and *invariants*. . .
- no polynomial-time algorithms are known for graph isomorphism.
- matrices equivalent under row and column operations iff equal rank; but tensor rank is NP-hard.
- derandomizing PIT implies circuit lower bounds [Kabanets-Impagliazzo]

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

Setup and orbit problems

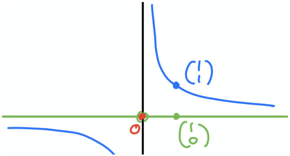
group $G \subseteq \mathrm{GL}_n(\mathbb{C})$ reductive, such as GL_n, SL_n, or $\mathrm{T}_n = (\mathbb{C}^*)^n$

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, but are there more efficient ways?

Setup and orbit problems

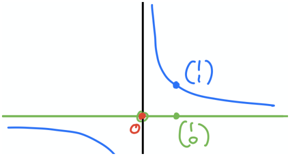
group $G \subseteq \mathrm{GL}_n(\mathbb{C})$ reductive, such as GL_n, SL_n, or $\mathrm{T}_n = (\mathbb{C}^*)^n$

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, but are there more efficient ways?

Setup and orbit problems

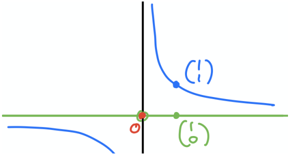
group $G \subseteq \mathrm{GL}_n(\mathbb{C})$ reductive, such as GL_n, SL_n, or $\mathrm{T}_n = (\mathbb{C}^*)^n$

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, but are there 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

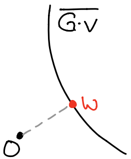
NB: The last two problems have a meaningful approximate version!

Null cone and norm minimization

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

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

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



$\|w\| = \min \{ \|u\| : u \in \overline{G \cdot v} \}$

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

Groups and derivatives

Thus 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 gradient?

Directional derivatives at $g = I$ are given by $\partial_{t=0} F_v(e^{At})$ for $A \in \mathrm{Lie}(G)$.

We may assume that maximal compact $K = G \cap \mathrm{U}_n$ acts by isometries.
Then we really optimize over $K \backslash G$, and it suffices to consider $A \in i\,\mathrm{Lie}(K)$.

For $G = \mathrm{GL}_n$: $\mathrm{U}_n \backslash \mathrm{GL}_n \cong \mathrm{PD}_n$ and $i\,\mathrm{Lie}(K) = \mathrm{Herm}_n$.

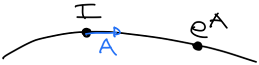
Groups and derivatives

Thus 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 gradient?

Directional derivatives at $g = I$ are given by $\partial_{t=0} F_v(e^{At})$ for $A \in \mathrm{Lie}(G)$.



We may assume that maximal compact $K = G \cap \mathrm{U}_n$ acts by isometries.
Then we really optimize over $K \backslash G$, and it suffices to consider $A \in i\,\mathrm{Lie}(K)$.

For $G = \mathrm{GL}_n$: $\mathrm{U}_n \backslash \mathrm{GL}_n \cong \mathrm{PD}_n$ and $i\,\mathrm{Lie}(K) = \mathrm{Herm}_n$.

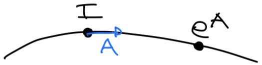
Groups and derivatives

Thus 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 gradient?

Directional derivatives at $g = I$ are given by $\partial_{t=0} F_v(e^{At})$ for $A \in \mathrm{Lie}(G)$.



We may assume that maximal compact $K = G \cap \mathrm{U}_n$ acts by isometries.
Then we really optimize over $K \backslash G$, and it suffices to consider $A \in i\,\mathrm{Lie}(K)$.

For $G = \mathrm{GL}_n$: $\mathrm{U}_n \backslash \mathrm{GL}_n \cong \mathrm{PD}_n$ and $i\,\mathrm{Lie}(K) = \mathrm{Herm}_n$.

Norm minimization and its dual

Thus we want to minimize the Kempf-Ness function:

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

The so-called moment map captures its *gradient* at $g = I$:

$$\mu \colon V \setminus \{0\} \to i\,\mathrm{Lie}(K), \quad \mathrm{tr}(\mu(v)H) = \partial_{t=0} F_v(e^{Ht}) \quad \forall H \in i\,\mathrm{Lie}(K)$$

- Clearly, $\mu(g \cdot v) = 0$ if g is minimizer.
- Remarkably, this is 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

Thus we want to minimize the Kempf-Ness function:

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

The so-called moment map captures its *gradient* at $g = I$:

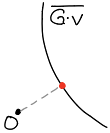
$$\mu \colon V \setminus \{0\} \to i\,\mathrm{Lie}(K), \quad \mathrm{tr}(\mu(v)H) = \partial_{t=0} F_v(e^{Ht}) \quad \forall H \in i\,\mathrm{Lie}(K)$$

- Clearly, $\mu(g \cdot v) = 0$ if g is minimizer.
- Remarkably, this is also sufficient! [Kempf-Ness]

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

Summary so far

$G \subseteq \mathrm{GL}_n$ complex reductive connected, $V = \mathbb{C}^m$ regular representation
$K = G \cap \mathrm{U}_n$ maximally compact, $\mu\colon V \setminus \{0\} \to i\,\mathrm{Lie}(K)$ 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. $\mu(g \cdot v) \approx 0$.

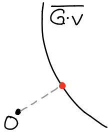
- The last two problems are dual, and either can solve null cone!
- But they also provide path to orbit closure intersection.

Useful *model problems*. Plausibly solvable in polynomial time, but rich enough to have interesting applications. Let us look at some. . .

Summary so far

$G \subseteq \mathrm{GL}_n$ complex reductive connected, $\quad V = \mathbb{C}^m$ regular representation
$K = G \cap \mathrm{U}_n$ maximally compact, $\quad \mu\colon V \setminus \{0\} \to i\,\mathrm{Lie}(K)$ moment map

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



...and its relaxations:

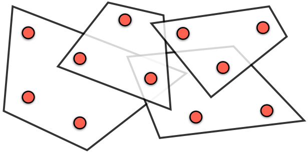
Norm minimization problem: Given v, find $g \in G$ s. th. $\|g \cdot v\| \approx \mathrm{cap}(v)$.

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

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

Useful *model problems*. Plausibly solvable in polynomial time, but 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

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

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 $\mathrm{conv}(S_n \cdot \lambda)$!

Kostka numbers: ... iff branching multiplicity for $\mathrm{T}_n \subset \mathrm{GL}_n$ is nonzero.

[Nonenmacher, 2008]

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

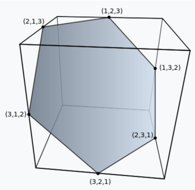
Example: Schur-Horn theorem

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]

Kostka numbers: . . . iff branching multiplicity for $T_n \subset GL_n$ is nonzero.

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

Example: Schur-Horn theorem

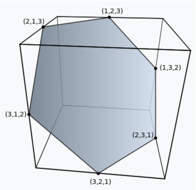
Let $\lambda_1 \geqslant \cdots \geqslant \lambda_n$ and $\delta_1, \ldots, \delta_n$ be integers.

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 $\operatorname{conv}(S_n \cdot \lambda)$!

Kostka numbers: . . . iff branching multiplicity for $T_n \subset GL_n$ is nonzero.



[Nonenmacher, 2008]

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

Torus actions

Let $T_n = (\mathbb{C}^*)^n$ act on $V = \bigoplus_{\omega \in \Omega} V_\omega$ with *weights* $\Omega \subseteq \mathbb{Z}^n$.
That is, if $v = \sum_\omega v_\omega$ then $z \cdot v = \sum_\omega z^\omega v_\omega$.

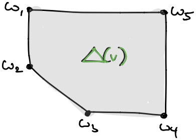
Capacity:

$$\operatorname{cap}(v)^2 = \inf_{z \in T_n} \sum_\omega |z^\omega|^2 \, \|v_\omega\|^2 = \inf_{x \in \mathbb{R}^n} \sum_\omega e^{x \cdot \omega} \, \|v_\omega\|^2$$

- norm minimization is geometric programming (log-convexity in x)
- $\operatorname{cap}(v) = 0$ iff $0 \notin \Delta(v) := \operatorname{conv}\{\omega : v_\omega \neq 0\}$; linear programming

Moment map:

$$\mu \colon V \setminus \{0\} \to \mathbb{R}^n, \quad \mu(v) = \frac{\sum_\omega \omega \|v_\omega\|^2}{\sum_\omega \|v_\omega\|^2}$$



- any point in $\Delta(v)$ can be approximately obtained [Atiyah]

Torus actions

Let $T_n = (\mathbb{C}^*)^n$ act on $V = \bigoplus_{\omega \in \Omega} V_\omega$ with *weights* $\Omega \subseteq \mathbb{Z}^n$.
That is, if $v = \sum_\omega v_\omega$ then $z \cdot v = \sum_\omega z^\omega v_\omega$.

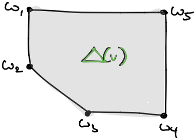
Capacity:

$$\operatorname{cap}(v)^2 = \inf_{z \in T_n} \sum_\omega |z^\omega|^2 \, \|v_\omega\|^2 = \inf_{x \in \mathbb{R}^n} \sum_\omega e^{x \cdot \omega} \, \|v_\omega\|^2$$

- norm minimization is geometric programming (log-convexity in x)
- $\operatorname{cap}(v) = 0$ iff $0 \notin \Delta(v) := \operatorname{conv}\{\omega : v_\omega \neq 0\}$; linear programming

Moment map:

$$\mu \colon V \setminus \{0\} \to \mathbb{R}^n, \quad \mu(v) = \frac{\sum_\omega \omega \|v_\omega\|^2}{\sum_\omega \|v_\omega\|^2}$$



- any point in $\Delta(v)$ can be approximately obtained [Atiyah]

Torus actions

Let $\mathrm{T}_n = (\mathbb{C}^*)^n$ act on $V = \bigoplus_{\omega \in \Omega} V_\omega$ with *weights* $\Omega \subseteq \mathbb{Z}^n$.
That is, if $v = \sum_\omega v_\omega$ then $z \cdot v = \sum_\omega z^\omega v_\omega$.

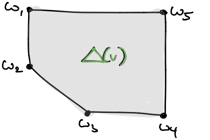
Capacity:

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

- norm minimization is geometric programming (log-convexity in x)
- $\operatorname{cap}(v) = 0$ iff $0 \notin \Delta(v) := \operatorname{conv}\{\omega : v_\omega \neq 0\}$; linear programming

Moment map:

$$\mu \colon V \setminus \{0\} \to \mathbb{R}^n, \quad \mu(v) = \frac{\sum_\omega \omega \|v_\omega\|^2}{\sum_\omega \|v_\omega\|^2}$$



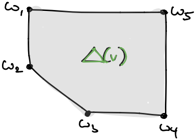
- any point in $\Delta(v)$ can be approximately obtained [Atiyah]

Moment polytopes

- For $G = T_n$, we saw on the previous slide that

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

is a convex polytope.



- For noncommutative G, get *magically* convex polytope. [Mumford, Kirwan, ...]
  E.g., for $G = GL_n$:

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

These are moment polytopes of G-orbit closures in $\mathbb{P}(V)$.

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

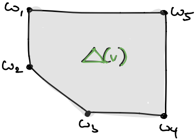
Even interesting when *not* restricting to orbits.

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 noncommutative G, get *magically* convex polytope. [Mumford, Kirwan, ...]
  E.g., for $G = \mathrm{GL}_n$:

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

These are moment polytopes of G-orbit closures in $\mathbb{P}(V)$.

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

Even interesting when *not* restricting to orbits.

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 $A + B = C$?

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

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

- count 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 in geometric complexity theory.

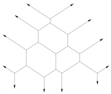
Example: Horn problem

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 $A + B = C$?

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

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



- count 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 in geometric complexity theory.

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

$$Y = (gX_1 h^{-1}, \ldots, gX_d h^{-1}) \qquad (g, h \in \mathrm{GL}_n)$$

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?

Polynomial identity testing: …iff $\exists$ matrices Y_k s.th. $\det \sum_k Y_k \otimes X_k \neq 0$.

- can solve in deterministic poly-time [Garg et al, cf. Ivanyos et al]
- when Y_k restricted to scalars: major open problem in TCS!

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?

Polynomial identity testing: . . . iff $\exists$ matrices Y_k s.th. $\det \sum_k Y_k \otimes X_k \neq 0$.

- can solve in deterministic poly-time [Garg et al, cf. Ivanyos et al]
- when Y_k restricted to scalars: major open problem in TCS!

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

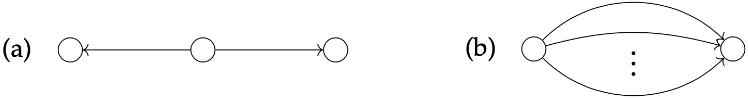
Example: Quivers

Quiver: Directed graph with vertex set Q_0 and edge set Q_1.

Given *dimension vector* $(n_x)_{x \in Q_0}$, consider natural action of

$$G = \prod_{x \in Q_0} \mathrm{GL}(n_x) \qquad \text{on} \qquad V = \bigoplus_{x \to y \in Q_1} \mathrm{Mat}_{n_y \times n_x}$$

- generalizes Horn and left-right action:



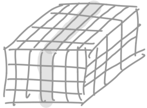
Many structural results known:

- semi-invariants characterized by [King, Derksen-Weyman, Schofield-Van den Bergh, ...]
- moment polytopes characterized by Horn-like inequalities [Baldoni-Vergne-W]

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?

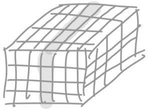
- eigenvalues form convex polytopes
- exponentially many vertices and faces
- characterized by asymptotic support of *Kronecker coefficients*
  NP-hard to determine if nonzero [Ikenmeyer-Mulmuley-W]

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

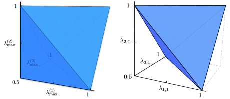
$$Y = (g_1 \otimes \ldots \otimes g_d) X \qquad (g_k \in \mathrm{GL}_{n_k})$$



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?

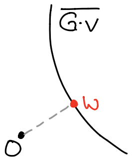


- eigenvalues form convex polytopes
- exponentially many vertices and faces
- characterized by asymptotic support of *Kronecker coefficients*

  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



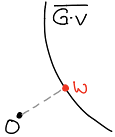
Non-commutative optimization duality

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:

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

[Kempf-Ness]



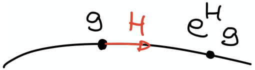
- primal: norm minimization, dual: scaling problem
- non-commutative version of linear programming duality

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

Why does the duality hold at all? F_v is convex along geodesics of $K \backslash G$! !

Geodesic convexity and smoothness

Homogeneous space $K \backslash G$ has geodesics $\gamma(t) = e^{tH} g$ for $H \in i\,\mathrm{Lie}(K)$.



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 $\|\mu(g \cdot v)\| \leqslant \varepsilon$ within $T = \frac{4N^2}{\varepsilon^2} \log \frac{\|v\|}{\mathrm{cap}(v)}$ iterations.

- Algorithm runs in time $\mathrm{poly}(\frac{1}{\varepsilon}, \text{input size})$.

  We use constructive invariant theory to give a priori lower bound on capacity.
- Algorithm solves null cone problem for suitable ε!

  Moment polytopes are rigid. We provide bound in terms of weight system.

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 $\|\mu(g \cdot v)\| \leqslant \varepsilon$ within $T = \frac{4N^2}{\varepsilon^2} \log \frac{\|v\|}{\mathrm{cap}(v)}$ iterations.

- Algorithm runs in time $\mathrm{poly}(\frac{1}{\varepsilon}, \text{input size})$.

  We use constructive invariant theory to give a priori lower bound on capacity.
- Algorithm solves null cone problem for suitable ε!

  Moment polytopes are rigid. We provide bound in terms of weight system.

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 to solve the null cone problem?

Theorem

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

To solve null cone problem, need two *a priori* lower bounds:

- Capacity bound: If $\mathrm{cap}(v) > 0$, then $\mathrm{cap}(v) \geqslant e^{-\mathrm{poly}(\text{input size})}$.
- Gradient bound: If $\mathrm{cap}(v) = 0$, then $\inf_{g \in G} \|\mu(g \cdot v)\| \geqslant \varepsilon_0$.

How to solve the null cone problem?

Theorem

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

To solve null cone problem, need two *a priori* lower bounds:

- Capacity bound: If $\operatorname{cap}(v) > 0$, then $\operatorname{cap}(v) \geqslant e^{-\operatorname{poly}(\text{input size})}$.

  Idea: Assume $v \in \mathbb{Z}^m$. Let p be G-invariant polynomial such that $p(v) \neq 0$. If p has degree D and integer coefficients bounded by L:

  $$1 \leqslant |p(v)| = |p(g \cdot v)| \leqslant m^D L \|g \cdot v\|^D \quad \Rightarrow \quad \|g \cdot v\| \geqslant \frac{1}{m L^{1/D}}.$$

  We can bound D and L using tools from invariant theory.

- Gradient bound: If $\operatorname{cap}(v) = 0$, then $\inf_{g \in G} \|\mu(g \cdot v)\| \geqslant \varepsilon_0$.

How to solve the null cone problem?

Theorem

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

To solve null cone problem, need two *a priori* lower bounds:

- Capacity bound: If $\operatorname{cap}(v) > 0$, then $\operatorname{cap}(v) \geqslant e^{-\operatorname{poly}(\text{input size})}$.
- Gradient bound: If $\operatorname{cap}(v) = 0$, then $\inf_{g \in G} \|\mu(g \cdot v)\| \geqslant \varepsilon_0$.

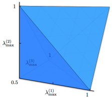
  Idea: There are finitely many possible moment polytopes $\Delta(v)$. Their facets are spannend by weights of the representation.

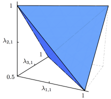
How about moment polytopes?

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:

- If $G = \mathrm{T}_n$ torus: simply shift weights $\omega \mapsto \omega - \lambda$
- If G noncommutative, more involved, need randomization [Mumford, Brion]

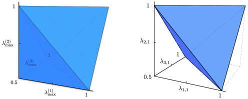
Result: Randomized first-order algorithm for moment polytopes.

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 $\lambda = 0$?

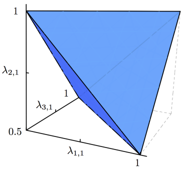


Shifting trick:

- If $G = \mathrm{T}_n$ torus: simply shift weights $\omega \mapsto \omega - \lambda$
- If G noncommutative, more involved, need randomization [Mumford, Brion]

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

Summary and outlook



$\boxed{\text{Null cone \& moment polytopes}}$

$\updownarrow$ duality

$\boxed{\text{Norm minimization}}$

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

- Polynomial-time algorithms in all cases?
- Better tools for geodesic optimization?
- Tensors in applications are often structured. Implications?
- What exponentially complex polytopes can be efficiently captured?
- What are the tractable problems in invariant theory? $\mathbb{C} \rightsquigarrow \mathbb{F}$?

Thank you for your attention!

A general equivalence

$\mathcal{V} \subseteq \mathbb{P}(V)$

All points in $\Delta(\mathcal{V})$ can be described via invariant theory:

$$V_\lambda \subseteq \mathbb{C}[\mathcal{V}]_{(k)} \quad \Rightarrow \quad \frac{\lambda}{k} \in \Delta(\mathcal{V})$$

(λ highest weight, k degree)

- Can also study multiplicities $g(\lambda, k) := \# V_\lambda \subseteq \mathbb{C}[\mathcal{V}]_{(k)}$.
- This leads to interesting computational problems:

$\boxed{g = ?}$	$\boxed{g > 0?}$	$\boxed{\exists s > 0 : g(s\lambda, sk) > 0?}$
(#-hard)	(NP-hard)	(our problem!)

Completely unlike Horn's problem: *Knutson-Tao saturation property does not hold, and hence we can hope for efficient algorithms!*