Invariants, polytopes, and optimization

Michael Walter

UNIVERSITY OF AMSTERDAM QuSoft Research Center for Quantum Software NWO

Workshop on Buildings, Varieties & Applications, Leipzig, November 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.

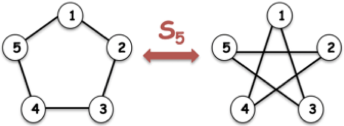
Plan for today:

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 determine when two objects are equivalent?

- computing *normal forms*, describing *moduli spaces* and *invariants*. . .

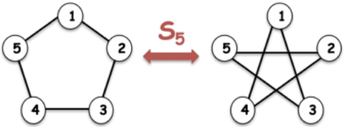
Interesting (and often difficult) problems with many applications:

- no polynomial-time algorithms are known for graph isomorphism.
- matrices equivalent under left-right action iff equal rank;
  but tensor rank is NP-hard.

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 determine when two objects are equivalent?

- computing *normal forms*, describing *moduli spaces* and *invariants*. . .

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.

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

General setup

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

$\pi \colon G \to \mathrm{GL}(V)$ regular representation on f.d. complex vector space

- orbits $Gv = \{\pi(g)v : g \in G\}$ and their closures $\overline{Gv}$

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 $\overline{Gv_1} \cap \overline{Gv_2} \neq \emptyset$?

- equivalently, $p(v_1) = p(v_2)$ for all G-invariant polynomials p
- captures equality in Mumford's *GIT quotient*

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

- v is called *unstable* if yes, *semistable* if no
- equivalently, $p(v) = p(0)$ for all G-invariant polynomials p

General setup

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

$\pi \colon G \to \mathrm{GL}(V)$ regular representation on f.d. complex vector space

- orbits $Gv = \{\pi(g)v : g \in G\}$ and their closures $\overline{Gv}$

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 $\overline{Gv_1} \cap \overline{Gv_2} \neq \emptyset$?

- equivalently, $p(v_1) = p(v_2)$ for all G-invariant polynomials p
- captures equality in Mumford's *GIT quotient*

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

- v is called *unstable* if yes, *semistable* if no
- equivalently, $p(v) = p(0)$ for all G-invariant polynomials p

Example: Conjugation

$G = \mathrm{GL}_n$, $V = \mathrm{Mat}_n$, $\pi(g)X = gXg^{-1}$

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

- X, Y are in *same orbit* iff same Jordan normal form
- X, Y have *intersecting orbit closures* iff same eigenvalues (counted with algebraic multiplicity)
- X is in *null cone* iff nilpotent

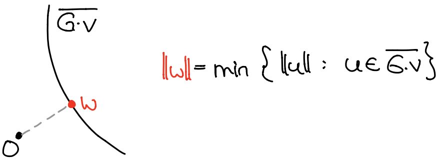
NB: The last two problems have a meaningful approximate version!

Null cone and norm minimization

We can characterize the null cone $\mathcal{N} = \{v \in V : 0 \in \overline{Gv}\}$ by an optimization problem. Capacity of v:

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

- v in null cone iff $\mathrm{cap}(v) = 0$



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

Norm minimization and its dual

Use K-invariant inner product, where $K = G \cap \mathrm{U}_n$ is maximal compact. We want to minimize the function:

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

Its gradient at $g = I$ defines the moment map:

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

(F_v should really be defined on $K \backslash G$; then $T_I \cong i\,\mathrm{Lie}(K)$; μ should be defined on $\mathbb{P}(V)$)

Kempf-Ness: Let $0 \neq w \in \overline{Gv}$. Then, $\mu(w) = 0$ iff w has minimal norm.

Thus we are led to:

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

Norm minimization and its dual

Use K-invariant inner product, where $K = G \cap \mathrm{U}_n$ is maximal compact.
We want to minimize the function:

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

Its gradient at $g = I$ defines the moment map:

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

(F_v should really be defined on $K \backslash G$; then $T_I \cong i\,\mathrm{Lie}(K)$; μ should be defined on $\mathbb{P}(V)$)

Kempf-Ness: Let $0 \neq w \in \overline{Gv}$. Then, $\mu(w) = 0$ iff w has minimal norm.

Thus we are led to:

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

Norm minimization and its dual

Use K-invariant inner product, where $K = G \cap \mathrm{U}_n$ is maximal compact. We want to minimize the function:

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

Its gradient at $g = I$ defines the moment map:

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

(F_v should really be defined on $K \backslash G$; then $T_I \cong i\,\mathrm{Lie}(K)$; μ should be defined on $\mathbb{P}(V)$)

Kempf-Ness: Let $0 \neq w \in \overline{Gv}$. Then, $\mu(w) = 0$ iff w has minimal norm.

Thus we are led to:

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

Summary so far

$G \subseteq \mathrm{GL}_n$ complex reductive, $\pi\colon G \to \mathrm{GL}(V)$ regular representation
$K \subseteq G$ maximally compact, $\mu\colon V \setminus \{0\} \to i\,\mathrm{Lie}(K)$ moment map

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

...and its relaxations:

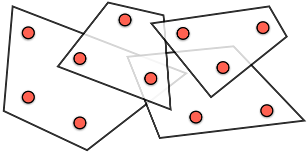
Norm minimization problem: Given v, find $g \in G$ s. th. $\|\pi(g)v\| \approx \mathrm{cap}(v)$.

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

The last two problems are dual to each other, and either can be used to solve null cone membership!

Let us look at some examples...

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

Permanent (Invariant Theory): ... 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 (Geometry): Given X, $\exists$ (approximately) d.s. scalings?

Permanent (Invariant Theory): ... 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 (Geometry): Given X, $\exists$ (approximately) d.s. scalings?

$$V = \mathrm{Mat}_n, \quad G = \mathrm{T}_n \times \mathrm{T}_n, \quad \pi(g,h)v = gvh.$$

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

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

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

Kostka numbers (Representation Theory): …iff branching multiplicity K_δ^λ for $\mathrm{T}_n \subset \mathrm{GL}_n$ is nonzero.

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

$V = V_\lambda$ Weyl module of GL_n, restricted to $G = T_n$.

Kc … K^λ_δ

for $T_n \subset \mathrm{GL}_n$ is nonzero.

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

Torus actions

Any representation of $G = T_n = (\mathbb{C}^*)^n$ decomposes as $V = \bigoplus_{\omega \in \Omega} V_\omega$ for *weights* $\Omega \subseteq \mathbb{Z}^n$. If $v = \sum_{\omega\in\Omega} v_\omega$ then $\pi(z)v = \sum_\omega z^\omega v_\omega$.

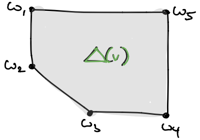
Capacity:

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

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

Moment map:

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



- Atiyah: $\overline{\mu(Gv)} = \Delta(v)$

Torus actions

Any representation of $G = T_n = (\mathbb{C}^*)^n$ decomposes as $V = \bigoplus_{\omega\in\Omega} V_\omega$ for *weights* $\Omega \subseteq \mathbb{Z}^n$. If $v = \sum_{\omega\in\Omega} v_\omega$ then $\pi(z)v = \sum_\omega z^\omega v_\omega$.

Capacity:

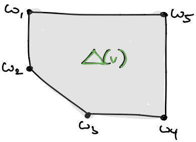
$$\operatorname{cap}(v)^2 = \inf_{z\in T_n} \sum_\omega |z^\omega|^2 \, \|v_\omega\|^2 = \inf_{x\in\mathbb{R}^n} \sum_\omega e^{x\cdot\omega} \, \|v_\omega\|^2$$

- 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:

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

- Atiyah: $\overline{\mu(Gv)} = \Delta(v)$



Torus actions

Any representation of $G = T_n = (\mathbb{C}^*)^n$ decomposes as $V = \bigoplus_{\omega \in \Omega} V_\omega$ for *weights* $\Omega \subseteq \mathbb{Z}^n$. If $v = \sum_{\omega \in \Omega} v_\omega$ then $\pi(z) v = \sum_\omega z^\omega v_\omega$.

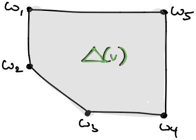
Capacity:

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

- 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:

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



- Atiyah: $\overline{\mu(Gv)} = \Delta(v)$

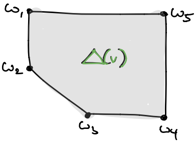
Moment polytopes

It is often interesting to characterize the image of the moment map:

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

$$\Delta(v) = \overline{\{\mu(w) : w \in Gv\}} \subseteq \mathbb{R}^n$$

is a convex polytope.



- If G non-commutative? For $G = \mathrm{GL}_n$, $\mu(w) \in \mathrm{Herm}_n$ and

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

is a convex polytope. General case similar. [Mumford, Kirwan]

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

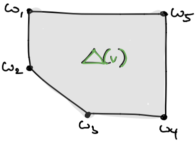
Moment polytope membership problem: Given v and λ, is $\lambda \in \Delta(v)$?

Often even interesting when *not* restricted to orbits. We will denote the corresponding polytope by Δ. It coincides with $\Delta(v)$ for generic v.

Moment polytopes

It is often interesting to characterize the image of the moment map:

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



$$\Delta(v) = \overline{\{\mu(w) : w \in Gv\}} \subseteq \mathbb{R}^n$$

is a convex polytope.

- If G non-commutative? For $G = \mathrm{GL}_n$, $\mu(w) \in \mathrm{Herm}_n$ and

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

is a convex polytope. General case similar. [Mumford, Kirwan]

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

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

Often even interesting when *not* restricted to orbits. We will denote the corresponding polytope by Δ. It coincides with $\Delta(v)$ for generic v.

Example: Horn problem

Let $\alpha_1 \geqslant \ldots \geqslant \alpha_n \geqslant 0$, $\beta_1 \geqslant \ldots \geqslant \beta_n \geqslant 0$, $\gamma_1 \geqslant \ldots \geqslant \gamma_n \geqslant 0$ be integers.

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

- Horn conjectured linear inequalities on α, β, γ.

Saturation property (Invariant theory): …iff *Littlewood-Richardson coefficient* $c^{\gamma}_{\alpha,\beta} > 0$ [Knutson-Tao]

- Horn inequalities sufficient
- lead to *only known* poly-time algorithm [Mulmuley]
- can find A, B, C by natural iterative algorithm [Franks]

Example: Horn problem

Let $\alpha_1 \geqslant \ldots \geqslant \alpha_n \geqslant 0$, $\beta_1 \geqslant \ldots \geqslant \beta_n \geqslant 0$, $\gamma_1 \geqslant \ldots \geqslant \gamma_n \geqslant 0$ be integers.

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

- Horn conjectured linear inequalities on α, β, γ.

Saturation property (Invariant theory): …iff *Littlewood-Richardson coefficient* $c^{\gamma}_{\alpha,\beta} > 0$ [Knutson-Tao]

- Horn inequalities sufficient
- lead to *only known* poly-time algorithm [Mulmuley]
- can find A, B, C by natural iterative algorithm [Franks]

Example: Horn problem

Let $\alpha_1 \geqslant \ldots \geqslant \alpha_n \geqslant 0$, $\beta_1 \geqslant \ldots \geqslant \beta_n \geqslant 0$, $\gamma_1 \geqslant \ldots \geqslant \gamma_n \geqslant 0$ be integers.

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

- Horn conjectured linear inequalities on α, β, γ.

Sa

co

ao]

ıley]

nks]

$$V = \mathrm{Mat}_n^2, \quad G = \mathrm{GL}_n \times \mathrm{GL}_n \times \mathrm{GL}_n,$$
$$\pi(g_1, g_2, g_3)(X, Y) := (g_1 X g_3^{-1}, g_2 Y g_3^{-1}).$$

$$\mu \colon V \setminus \{0\} \to \mathrm{Herm}_n^3$$
$$\mu(X, Y) = (XX^*, YY^*, -X^*X - Y^*Y)$$

$$\Delta = \{(\alpha, \beta, -\gamma) \ : \ A \geqslant 0, B \geqslant 0, \mathrm{tr}(A) + \mathrm{tr}(B) = 1\}$$

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 (q.d.s.)* if $\sum_k X_k X_k^* = \sum_k X_k^* X_k = I$.

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

Polynomial identity testing (Invariant Theory): ... iff $\exists$ matrices Y_k such that $\sum_k Y_k \otimes X_k$ is invertible.

- numerical algorithms can solve this in deterministic polynomial 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 (q.d.s.)* if $\sum_k X_k X_k^* = \sum_k X_k^* X_k = I$.

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

Polynomial identity testing (Invariant Theory): ...iff $\exists$ matrices Y_k such that $\sum_k Y_k \otimes X_k$ is invertible.

- numerical algorithms can solve this in deterministic polynomial 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 (q.d.s.)* if $\sum_k X_k X_k^* = \sum_k X_k^* X_k = I$.

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

$V = \mathrm{Mat}_n^d, \quad G = \mathrm{GL}_n \times \mathrm{GL}_n, \quad \pi(g,h)$ as above.

$\mu \colon V \setminus \{0\} \to \mathrm{Herm}_n \oplus \mathrm{Herm}_n$

$\mu(X_1, \ldots, X_d) = (\sum_k X_k X_k^*, -\sum_k X_k^* X_k)$

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:

(a) (b)

Many structural results known:

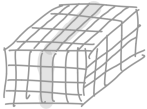
- semi-invariants characterized by [King, Derksen-Weyman, Schofield-Van den Bergh, . . .]
- moment polytopes characterized by Horn-like inequalities [Baldoni-Vergne-W]

. . . but efficient algorithms only in special cases.

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 principal flattening of X.

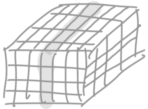
(In quantum mechanics, X describes joint state of d particles and ρ_k marginal of k-th particle.)

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

- eigenvalues form convex polytopes (moment polytopes)
- exponentially many vertices, faces [Berenstein-Sjamaar, Klyachko, Ressayre, Vergne-W]
- related to asymptotic support of *Kronecker coefficients*
- 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

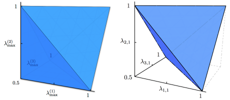


$$Y = (g_1 \otimes \ldots \otimes g_d) X \qquad (g_k \in \mathrm{GL}_{n_k})$$

Consider $\rho_k = X_k X_k^*$, where X_k is k-th principal 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?

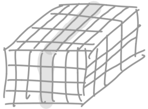


- eigenvalues form convex polytopes (moment polytopes)
- exponentially many vertices, faces [Berenstein-Sjamaar, Klyachko, Ressayre, Vergne-W]
- related to asymptotic support of *Kronecker coefficients*
- can we find efficient algorithmic description?

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

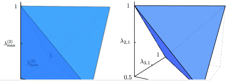
$$Y = (g_1 \otimes \ldots \otimes g_d)X \qquad (g_k \in \mathrm{GL}_{n_k})$$



Consider $\rho_k = X_k X_k^*$, where X_k is k-th principal 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?



$V = \mathbb{C}^{n_1} \otimes \ldots \otimes \mathbb{C}^{n_d}$, $\quad G = \mathrm{GL}_{n_1} \times \cdots \times \mathrm{GL}_{n_d}$, $\quad \pi$ as above.

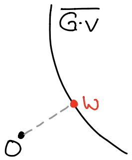
$\mu \colon V \setminus \{0\} \to \mathrm{Herm}_{n_1} \oplus \ldots \oplus \mathrm{Herm}_{n_d}$

$\mu(v) = (\rho_1, \ldots, \rho_d)$

$\Delta(v) = \{(\operatorname{spec} \rho_1, \ldots, \operatorname{spec} \rho_d)\}$

-W]

Geodesic first-order algorithms for norm minimization and scaling

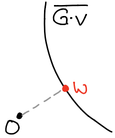


Non-commutative optimization duality

Recall $F_v(g) = \log\|\pi(g)v\|$ and $\mu(v)$ is its gradient at $g = I$. By Kempf-Ness, the following *optimization problems* are equivalent:

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

[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: $\partial^2_{t=0} F_v(\gamma(t)) \leqslant 2N(\pi)^2 \|H\|^2$

$N(\pi)$ is the *weight norm*, defined as the maximal norm of all weights in π.

- typically small (e.g., upper-bounded by degree for $G = \mathrm{GL}_n$)

Smoothness implies that

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

Thus, gradient descent with sufficiently small step size makes progress!

First-order algorithm: geodesic gradient descent

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

Algorithm: Start with $g = I$. For $t = 1, \ldots, T$:
Compute moment map $\mu(w)$ of $w = \pi(g)v$. 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 $T = \frac{4N(\pi)^2}{\varepsilon^2} \log \frac{\|v\|}{\operatorname{cap}(v)}$ iterations.

- Algorithm runs in time $\operatorname{poly}(\frac{1}{\varepsilon}, \text{input size})$.
  Can use constructive invariant theory to lower-bound capacity.
- Algorithm solves null cone membership problem if ε sufficiently small!
  Moment polytopes are rigid thanks to geometric invariant theory.

Peter Bürgisser will explain this in more detail tomorrow.

First-order algorithm: geodesic gradient descent

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

Algorithm: Start with $g = I$. For $t = 1, \ldots, T$:
Compute moment map $\mu(w)$ of $w = \pi(g)v$. 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 $T = \frac{4N(\pi)^2}{\varepsilon^2} \log \frac{\|v\|}{\mathrm{cap}(v)}$ iterations.

- Algorithm runs in time $\mathrm{poly}(\frac{1}{\varepsilon}, \text{input size})$.
  Can use constructive invariant theory to lower-bound capacity.
- Algorithm solves null cone membership problem if ε sufficiently small!
  Moment polytopes are rigid thanks to geometric invariant theory.

Peter Bürgisser will explain this in more detail tomorrow.

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:

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

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

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

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

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

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:

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

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

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

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

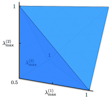
$$F_v(g_{\text{new}}) \leqslant F_v(g) - \frac{1}{4N(\pi)^2}\|\mu(w)\|^2 \leqslant F_v(g) - \frac{\varepsilon^2}{4N(\pi)^2}. \qquad \square$$

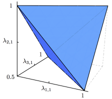
How about moment polytopes?

Recall:

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

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





Shifting trick:

- for simplicity, assume λ integral
- replace V by $W = V \otimes V_{\lambda^*}$ if G commutative, shifts all weights by $-\lambda$
- $\lambda \in \Delta(v)$ iff $0 \in \Delta(w)$ for *generic* $w \in v \otimes \pi(G) v_{\lambda^*}$ ☺   [Mumford, Brion, ...]

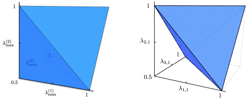
Result: Randomized first-order algorithm for moment polytopes.

How about moment polytopes?

Recall:

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

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

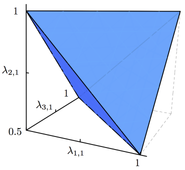


Shifting trick:

- for simplicity, assume λ integral
- replace V by $W = V \otimes V_{\lambda^*}$ — if G commutative, shifts all weights by $-\lambda$
- $\lambda \in \Delta(v)$ iff $0 \in \Delta(w)$ for *generic* $w \in v \otimes \pi(G) v_{\lambda^*}$ ☺ [Mumford, Brion, ...]

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

Summary and outlook



Null cone & moment polytopes

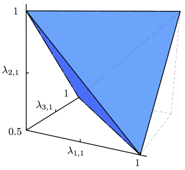
$\updownarrow$ duality

Norm minimization

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

On Tuesday, 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 geometric invariant theory and geodesic convex optimization, with a wide range of applications. *Many exciting directions:*

- Polynomial-time algorithms in all cases?
- In commutative case, poly-time algorithms known and can beat our geodesic algorithms! Can we design geodesic interior point methods?
- Tensors in applications are often structured. Implications?
- 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:

$g = ?$	$g > 0?$	$\exists s > 0 : g(s\lambda, sk) > 0?$
(#-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!*