Invariants, polytopes, and optimization

Michael Walter

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 $\leftrightarrow$ 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: lower bounds from *symmetry obstructions* [Mulmuley]
- 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...
Symmetries and group actions

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

![Diagram showing group action](image)

**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*.
- derandomizing *PIT* implies circuit lower bounds

We will see many more examples in a moment...
Setup and orbit problems

$G \subseteq \text{GL}_n(\mathbb{C})$ group such as $\text{GL}_n$, $\text{SL}_n$, or $T_n =$ invertible diagonal matrices

$\pi: G \to \text{GL}(V)$ representation on vector space $V = \mathbb{C}^m$

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

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

$$\pi(g)\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} gx \\ g^{-1}y \end{pmatrix}$$

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

\( G \subseteq \text{GL}_n(\mathbb{C}) \) group such as \( \text{GL}_n, \text{SL}_n, \) or \( T_n = \) invertible diagonal matrices

\( \pi: G \to \text{GL}(V) \) representation on vector space \( V = \mathbb{C}^m \)

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

Example: \( G = \text{GL}_1 = \mathbb{C}^*, \: V = \mathbb{C}^2 \)

\[ \pi(g) \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} gx \\ g^{-1}y \end{pmatrix} \]

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

$G \subseteq \text{GL}_n(\mathbb{C})$ group such as $\text{GL}_n$, $\text{SL}_n$, or $T_n = \text{invertible diagonal matrices}$

 Representation $\pi: G \to \text{GL}(V)$ on vector space $V = \mathbb{C}^m$

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

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

$$\pi(g) \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} gx \\ g^{-1}y \end{pmatrix}$$

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

\[ G = \text{GL}_n, \quad V = \text{Mat}_n, \quad \pi(g)X = gXg^{-1} \]

\[
\begin{pmatrix}
\lambda_1 & 1 \\
\lambda_1 & 1 \\
\lambda_1 & \ddots
\end{pmatrix}
\]

- \(X, Y\) 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 translate the null cone problem into an optimization problem. Define capacity of $v$:

$$\text{cap}(v) := \min_{u \in \mathbb{G}v} \|u\| = \inf_{g \in G} \|\pi(g)v\|$$

- clearly, $0 \in \mathbb{G}v$ iff $\text{cap}(v) = 0$
- generalizes Gurvits’ notions of matrix, polynomial, operator capacity

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

We want to minimize the function:

\[ F_v : G \rightarrow \mathbb{R}, \quad F_v(g) := \log \| \pi(g)v \| \]

First-order condition? How to define derivatives?

Consider \( G = \text{GL}_n \). Any invertible matrix \( g \) can be written as exponential:

\[ \text{GL}_n = \{ g = e^A : A \in \text{Mat}_n \} \]

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 \( G \subseteq \text{GL}_n \) – only need to restrict allowed directions.
Groups and derivatives

We want to minimize the function:

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

**First-order condition? How to define derivatives?**

Consider \( G = \text{GL}_n \). Any invertible matrix \( g \) can be written as exponential:

\[ \text{GL}_n = \{ g = e^A : A \in \text{Mat}_n \} \]

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 \( G \subseteq \text{GL}_n \) – only need to restrict allowed directions.
Norm minimization and its dual

We want to minimize the function:

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

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:

\[ \mu : V \setminus \{0\} \to \text{Herm}_n, \quad \text{tr}(\mu(v)A) = \partial_{t=0} F_v(e^{At}) \quad \forall A \]

- clearly, \( \mu(\pi(g)v) = 0 \) if \( g \) is minimizer
- amazingly, also sufficient

Scaling problem: Given \( v \), find \( g \in G \) such that \( \mu(\pi(g)v) \approx 0 \). [Kempf-Ness]
Norm minimization and its dual

We want to minimize the function:

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

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:

\[ \mu : V \setminus \{0\} \to \text{Herm}_n, \quad \text{tr}(\mu(v)A) = \partial_{t=0} F_v(e^{At}) \quad \forall A \]

- clearly, \( \mu(\pi(g)v) = 0 \) if \( g \) is minimizer
- amazingly, also sufficient

Scaling problem: Given \( v \), find \( g \in G \) such that \( \mu(\pi(g)v) \approx 0 \).
Norm minimization and its dual

We want to minimize the function:

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

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:

\[ \mu : V \setminus \{0\} \to \text{Herm}_n, \quad \text{tr}(\mu(v)A) = \partial_{t=0} F_v(e^{At}) \quad \forall A \]

- clearly, \( \mu(\pi(g)v) = 0 \) if \( g \) is minimizer
- amazingly, also sufficient

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

$G \subseteq \text{GL}_n$ group, $\pi: G \to \text{GL}(V)$ rep., $\mu: V \setminus \{0\} \to \text{Herm}_n$ 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. $\|\pi(g)v\| \approx \text{cap}(v)$.

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!
▶ 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

$G \subseteq \text{GL}_n$ group, $\pi: G \to \text{GL}(V)$ rep., $\mu: V \setminus \{0\} \to \text{Herm}_n$ 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. $\|\pi(g)v\| \approx \text{cap}(v)$.

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!
- 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 a matrix with nonnegative entries. A scaling of $X$ is a matrix

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

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

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

**Permanent:** $\ldots$ iff $\text{per}(X) > 0!$

- $\ldots$ iff $\exists$ bipartite perfect matching in support of $X$
- can be decided in polynomial time
- find scalings by alternatingly fixing rows & columns
- convergence controlled by permanent

Connections to statistics, complexity, combinatorics, geometry, numerics, …
Example: Matrix scaling (raking, IPFP, . . .)

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

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

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 $\text{per}(X) > 0$!

- . . . 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 a matrix with nonnegative entries. A scaling of $X$ is a matrix

$$Y = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} X \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$$

$(a_1, \ldots, b_n > 0)$.

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

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

**Permanent:** $\per(X) > 0$ if and only if there is a perfect matching in the 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: Schur-Horn theorem

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

Given $\lambda$ and $\delta$, $\exists$ Hermitian matrix with spectrum $\lambda$ and diagonal $\delta$?

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

Schur-Horn theorem: $\ldots$ iff $\delta$ in ‘permutahedron’ generated by $\lambda$, 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, \ldots]
Example: Schur-Horn theorem

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

Given $\lambda$ and $\delta$, $\exists$ Hermitian matrix with spectrum $\lambda$ and diagonal $\delta$?

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

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

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

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

Given $\lambda$ and $\delta$, $\exists$ Hermitian matrix with spectrum $\lambda$ and diagonal $\delta$?

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

Schur-Horn theorem: $\delta$ is aff. to $\lambda$ iff $\delta$ is in 'permutahedron' generated by $\lambda$, 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, ...]

$V = V_\lambda$ irreducible representation of $GL_n$, restricted to $G = T_n$. 

[Nonenmacher, 2008]
Example: Laurent polynomials

The group $T_n$ acts on Laurent polynomials $P = \sum_{\omega} p_{\omega} Z_1^{\omega_1} \cdots Z_n^{\omega_n}$ by scaling variables: $\pi(g)P = \sum_{\omega} p_{\omega}(g_1 Z_1)^{\omega_1} \cdots (g_n Z_n)^{\omega_n}$.

Capacity:

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

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

Moment map:

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

- any point in $\Delta(P)$ can be approximated by scalings of $P$ [Atiyah]
Example: Laurent polynomials

The group $T_n$ acts on Laurent polynomials $P = \sum_\omega p_\omega Z_1^{\omega_1} \cdots Z_n^{\omega_n}$ by scaling variables: $\pi(g)P = \sum_\omega p_\omega (g_1 Z_1)^{\omega_1} \cdots (g_n Z_n)^{\omega_n}$.

**Capacity:**

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

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

**Moment map:**

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

- any point in $\Delta(P)$ can be approximated by scalings of $P$ [Atiyah]

![Newton polytope](image-url)
Example: Laurent polynomials

The group $T_n$ acts on Laurent polynomials $P = \sum_\omega p_\omega Z^{\omega_1} \cdots Z^{\omega_n}$ by scaling variables: $\pi(g)P = \sum_\omega p_\omega (g_1 Z)^{\omega_1} \cdots (g_n Z)^{\omega_n}$.

**Capacity:**

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

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

**Moment map:**

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

- any point in $\Delta(P)$ can be approximated by scalings of $P$ [Atiyah]
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
  \[ \Delta(v) = \{ \mu(w) : w \in Gv \} \subseteq \mathbb{R}^n \]
  is a convex polytope.

- If $G$ non-commutative? For $G = GL_n$, $\mu(w) \in \text{Herm}_n$ and
  \[ \Delta(v) = \{ \text{spec}(\mu(w)) : w \in Gv \} \subset \mathbb{R}^n \]
  is (rather magically) a convex polytope.

These polytopes are known as moment polytopes.

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

Often even interesting when not restricted to orbit of a single vector $v$. We will denote the corresponding polytope by $\Delta$. 

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
  \[ \Delta(v) = \{ \mu(w) : w \in Gv \} \subseteq \mathbb{R}^n \]
  is a convex polytope.

- If $G$ non-commutative? For $G = \text{GL}_n$, $\mu(w) \in \text{Herm}_n$ and
  \[ \Delta(v) = \{ \text{spec}(\mu(w)) : w \in Gv \} \subset \mathbb{R}^n \]
  is (rather magically) a convex polytope.

These polytopes are known as moment polytopes.

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

Often even interesting when not restricted to orbit of a single vector $v$. We will denote the corresponding polytope by $\Delta$. 
Example: Horn problem

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

Horn problem: When $\exists$ Hermitian $n \times n$ matrices $A$, $B$, $C$ with spectrum $\alpha$, $\beta$, $\gamma$ such that $A + B = C$?

- exponentially many linear inequalities on $\alpha$, $\beta$, $\gamma$  
- e.g., $\alpha_1 + \beta_1 \geq \gamma_1$

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 for Horn problem

- can find $A$, $B$, $C$ by natural (yet inefficient) algorithm

Motivation for Mulmuley's positivity hypotheses.
Example: Horn problem

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

**Horn problem:** When $\exists$ Hermitian $n \times n$ matrices $A$, $B$, $C$ with spectrum $\alpha$, $\beta$, $\gamma$ such that $A + B = C$?

- exponentially many linear inequalities on $\alpha$, $\beta$, $\gamma$ [Horn]
- e.g., $\alpha_1 + \beta_1 \geq \gamma_1$

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

- counts multiplicities in representation theory, combinatorial gadgets, integer points in polytopes, $\ldots$
- poly-time algorithm for Horn problem
- can find $A$, $B$, $C$ by natural (yet inefficient) algorithm

Motivation for Mulmuley’s positivity hypotheses.
Example: Horn problem

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

**Horn problem:** When $\exists$ Hermitian $n \times n$ matrices $A$, $B$, $C$ with spectrum $\alpha$, $\beta$, $\gamma$ such that $A + B = C$?

- exponentially many linear inequalities on $\alpha$, $\beta$, $\gamma$
  
  - e.g., $\alpha_1 + \beta_1 \geq \gamma_1$

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

$$\mu: V \setminus \{0\} \rightarrow \text{Herm}_n^3$$

$$\mu(X, Y) = (XX^*, YY^*, -X^*X - Y^*Y)$$

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

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

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

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: $\ldots$ iff $\exists$ matrices $Y_k$ s. th. $\sum_k Y_k \otimes X_k$ invertible.

- can solve in deterministic poly-time $\text{[Garg et al, cf. Ivanyos et al]}$
- when $Y_k$ restricted to scalars: PIT for symbolic determinants $\n 

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

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

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: $\ldots$ 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, $\ldots$).
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

$$Y = (gX_1 h^{-1}, \ldots, gX_d h^{-1})$$

$(g, h \in \text{GL}_n)$

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?

Notation:

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

$$\mu: V \setminus \{0\} \to \text{Herm}_n \oplus \text{Herm}_n$$

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

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

$$Y = (g_1 \otimes \cdots \otimes g_d)X \quad (g_k \in \text{GL}_{n_k})$$

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 $\rho_k$ 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 with exponentially many vertices and faces
  [Berenstein-Sjamaar, Klyachko, Ressayre, Vergne-W]
- related to asymptotics of Kronecker coefficients
  NP-hard to determine if nonzero
  [Ikenmeyer-Mulmuley-W]
- can we find efficient algorithmic description? key challenge!
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

$$Y = (g_1 \otimes \cdots \otimes g_d)X \quad (g_k \in \text{GL}_{n_k})$$

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 $\rho_k$ 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 with exponentially many vertices and faces
  [Berenstein-Sjamaar, Klyachko, Ressayre, Vergne-W]

- related to asymptotics of Kronecker coefficients
  NP-hard to determine if nonzero
  [Ikenmeyer-Mulmuley-W]

- can we find efficient algorithmic description? key challenge!
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

$$Y = (g_1 \otimes \cdots \otimes g_d)X \quad (g_k \in \text{GL}_{n_k})$$

Consider $\rho_k = X_kX_k^*$, where $X_k$ is $k$-th flattening of $X$. (In quantum mechanics, $X$ describes joint state of $d$ particles and $\rho_k$ marginal of $k$-th particle.)

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

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

$\mu: V \setminus \{0\} \rightarrow \text{Herm}_{n_1} \oplus \cdots \oplus \text{Herm}_{n_d}$

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

$\Delta(v) = \{(\text{spec } \rho_1, \ldots, \text{spec } \rho_d)\}$
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 $g = I$. We discussed that the following optimization problems are equivalent:

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

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

**Proposition:** $F_v$ satisfies the following properties along these geodesics:

1. **Convexity:** $\partial^2_{t=0} F_v(\gamma(t)) \geq 0$
2. **Smoothness:** $\partial^2_{t=0} F_v(\gamma(t)) \leq 2N^2\|H\|^2$

$N = N(\pi)$ is a small constant, upper-bounded by degree.

**Smoothness** implies that

$$F_v(e^Hg) \leq F_v(g) + \text{tr}(\mu(v)H) + N^2\|H\|^2.$$

Thus, gradient descent makes progress if steps not too large!
First-order algorithm: geodesic gradient descent

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

**Algorithm:** Start with \( g = I \). For \( t = 1, \ldots, T \):

1. Compute moment map \( \mu(w) \) of \( w = \pi(g)v \). If norm \( \varepsilon \)-small, stop.
2. Otherwise, replace \( g \) by \( e^{-\eta \mu(w)}g \). \( \eta > 0 \) suitable step size

**Theorem**

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

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

Peter Bürgisser will explain this in more detail.
First-order algorithm: geodesic gradient descent

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

**Algorithm:** Start with \( g = I \). For \( t = 1, \ldots, T \):

- Compute moment map \( \mu(w) \) of \( w = \pi(g)v \). If norm \( \varepsilon \)-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 \( \text{cap}(v) > 0 \). Then the algorithm outputs \( g \in G \) such that \( \|\mu(w)\| \leq \varepsilon \) within \( T = \frac{4N^2}{\varepsilon^2} \log \frac{\|v\|}{\text{cap}(v)} \) iterations.

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

Peter Bürgisser will explain this in more detail.
Analysis of algorithm

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

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

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

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

Progress follows from smoothness:

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

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

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

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

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

Analysis of algorithm

“Unless moment map \( \varepsilon \)-small, replace \( g \) by \( e^{-\eta \mu(w)} g \).”

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

\[
F_v(g_{\text{new}}) \leq F_v(g) - c
\]

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

Progress follows from smoothness:

\[
F_v(e^H g) \leq F_v(g) + \text{tr}(\mu(v) H) + N^2 \|H\|^2
\]

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

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

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

\[
F_v(g_{\text{new}}) \leq F_v(g) - \frac{1}{4N^2} \|\mu(w)\|^2 \leq F_v(g) - \frac{\varepsilon^2}{4N^2}.
\]

\( \square \)
How about moment polytopes?

Recall:

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

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

**Shifting trick**:

- if $G$ commutative, simply shifts polytope $\omega \mapsto \omega - \lambda$
- if $G$ noncommutative, more involved and uses randomization

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

Recall:

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

- $\nu$ in null cone $\iff 0 \not\in \Delta(\nu)$
- how to reduce to $\lambda = 0$?

**Shifting trick:**

- if $G$ commutative, simply shifts polytope $\omega \mapsto \omega - \lambda$
- if $G$ noncommutative, more involved and uses randomization

**Result:** *Randomized* first-order algorithm for moment polytopes.
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.
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? \( \mathbb{C} \sim \mathbb{F} \)?

Thank you for your attention!