# Noncommutative Group Symmetries and Optimization

Michael Walter (University of Amsterdam)

based on joint works with Peter Bürgisser, Cole Franks, Ankit Garg,
Rafael Oliveira, Avi Wigderson

Workshop on Geometry and Optimization in Quantum Information,
Oberwolfach, October 2021

# Prelude: Matrix scaling

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 problem: Given $X$, find (approximately) d.s. scalings.

Algebra: Possible iff $\operatorname{per}(X) > 0$!

- . . . iff $\exists$ bipartite perfect matching in support of $X$
- can be decided in polynomial time

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

# Prelude: Matrix scaling

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 problem: Given $X$, find (approximately) d.s. scalings.

Algebra: Possible iff $\operatorname{per}(X) > 0$!

- ... iff $\exists$ bipartite perfect matching in support of $X$
- can be decided in polynomial time

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

# Sinkhorn (flip flop) algorithm

To scale matrix, alternatingly normalize rows & columns:

$$\begin{pmatrix} 1 & 2 \\ 4 & 0 \end{pmatrix} \xrightarrow{\text{rows}} \begin{pmatrix} 1/3 & 2/3 \\ 1 & 0 \end{pmatrix} \xrightarrow{\text{cols}} \begin{pmatrix} 1/4 & 1 \\ 3/4 & 0 \end{pmatrix} \longrightarrow \dots \longrightarrow \begin{pmatrix} \varepsilon & 1 \\ 1-\varepsilon & 0 \end{pmatrix}$$

This converges whenever possible! In turn, scaled matrices give exponential approximations for permanent.

Three questions:

- Why does such a simple “greedy” algorithm work?
- What is the connection between scaling and the permanent?
- Is there a general perspective?

# Sinkhorn (flip flop) algorithm

To scale matrix, alternatingly normalize rows & columns:

$$\begin{pmatrix} 1 & 2 \\ 4 & 0 \end{pmatrix} \xrightarrow{\text{rows}} \begin{pmatrix} 1/3 & 2/3 \\ 1 & 0 \end{pmatrix} \xrightarrow{\text{cols}} \begin{pmatrix} 1/4 & 1 \\ 3/4 & 0 \end{pmatrix} \longrightarrow \dots \longrightarrow \begin{pmatrix} \varepsilon & 1 \\ 1-\varepsilon & 0 \end{pmatrix}$$

This converges whenever possible! In turn, scaled matrices give exponential approximations for permanent.

Three questions:

- Why does such a simple “greedy” algorithm work?
- What is the connection between scaling and the permanent?
- Is there a general perspective?

# Overview

There are geometric and algebraic problems, arising from group actions, that are amenable to geodesic convex optimization.

$$\boxed{\text{Scaling \& marginal problems}} \quad \longleftrightarrow \quad \boxed{\text{Norm minimization}}$$

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

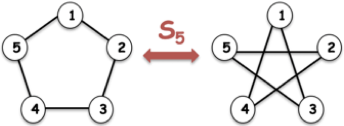
Plan for today:

1. Introduction to the framework
2. Panorama of applications
3. Geometry
4. Algorithmic solutions

*Optimization algorithms for problems with natural symmetries!*

# Symmetries and group actions

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



Problem: How can we algorithmically and efficiently check equivalence?

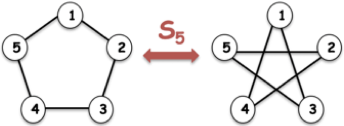
Interesting (and often difficult) problems with many applications:

- no polynomial-time algorithm known for graph isomorphism
- matrices equivalent under row and column operations iff equal rank; not true for tensor rank which is also NP-hard
- derandomizing polynomial identity testing implies circuit lower bounds
- computing *normal forms*, describing *moduli spaces* and *invariants*...

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?

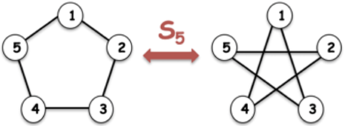
Interesting (and often difficult) problems with many applications:

- no polynomial-time algorithm known for graph isomorphism
- matrices equivalent under row and column operations iff equal rank; not true for tensor rank which is also NP-hard
- derandomizing polynomial identity testing implies circuit lower bounds
- computing *normal forms*, describing *moduli spaces* and *invariants*. . .

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?

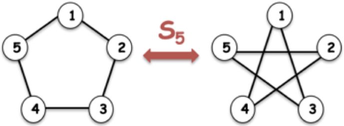
Interesting (and often difficult) problems with many applications:

- no polynomial-time algorithm known for graph isomorphism
- matrices equivalent under row and column operations iff equal rank; not true for tensor rank which is also NP-hard
- derandomizing polynomial identity testing implies circuit lower bounds
- computing *normal forms*, describing *moduli spaces* and *invariants*. . .

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 algorithm known for graph isomorphism
- matrices equivalent under row and column operations iff equal rank; not true for tensor rank which is also NP-hard
- derandomizing polynomial identity testing implies circuit lower bounds
- computing *normal forms*, describing *moduli spaces* and *invariants*. . .

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

# Orbit problems

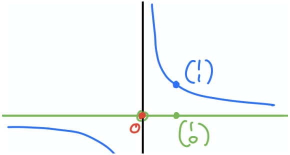
**Group** $G \subseteq \mathrm{GL}_n(\mathbb{C})$ “nice” (reductive), such as $\mathrm{GL}_n$, $\mathrm{SL}_n$, or $\mathrm{T}_n = (\mathbb{C}^*)^n$

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

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

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

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



**Orbit problems:**

- Given $v$ and $w$, are they in the same orbit? That is, is $Gv = Gw$?
- Robust versions: $w \in \overline{Gv}$? $\quad \overline{Gv} \cap \overline{Gw} \neq \emptyset$?
- **Null cone problem:** $0 \in \overline{Gv}$?

# Orbit problems

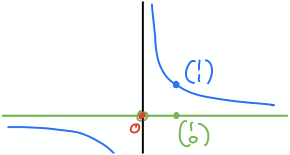
**Group** $G \subseteq \mathrm{GL}_n(\mathbb{C})$ “nice” (reductive), such as $\mathrm{GL}_n$, $\mathrm{SL}_n$, or $\mathrm{T}_n = (\mathbb{C}^*)^n$

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

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

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

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



**Orbit problems:**

- Given $v$ and $w$, are they in the same orbit? That is, is $Gv = Gw$?
- Robust versions: $w \in \overline{Gv}$? $\quad \overline{Gv} \cap \overline{Gw} \neq \emptyset$?
- **Null cone problem:** $0 \in \overline{Gv}$?

# Example: Conjugation

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

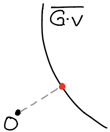
$$\begin{pmatrix} \lambda_1 & 1 & & \\ & \lambda_1 & & \\ & & \lambda_2 & \\ & & & \ddots \end{pmatrix}$$

- $X$, $Y$ 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

# Big picture: Null cone, optimization, and scaling

For concreteness, focus on **null cone problem**:

Is $0 \in \overline{Gv}$?



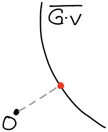
# Big picture: Null cone, optimization, and scaling

Is $P(v) = P(0)$ for every *invariant* polynomial $P$?

Algebra

$$\Updownarrow$$

Is $0 \in \overline{Gv}$?

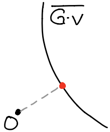


# Big picture: Null cone, optimization, and scaling

Is $P(v) = P(0)$ for every *invariant* polynomial $P$? Algebra

$\Updownarrow$

Is $0 \in \overline{Gv}$?

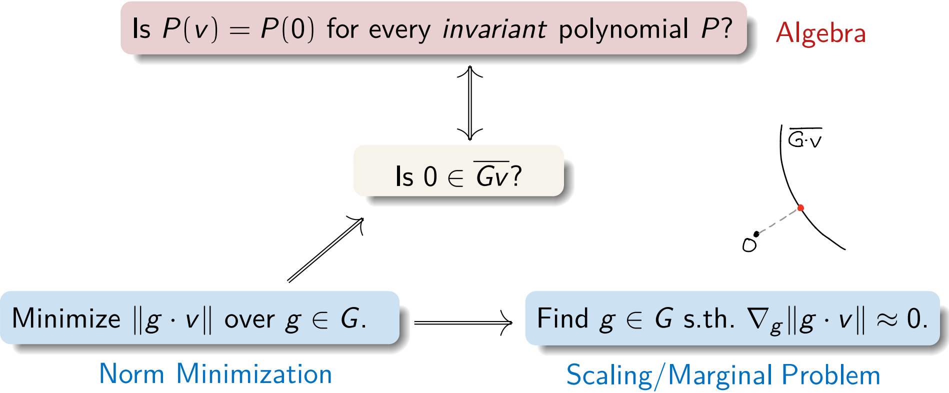


$\nearrow$

Minimize $\|g \cdot v\|$ over $g \in G$.

Norm Minimization

# Big picture: Null cone, optimization, and scaling



Is $P(v) = P(0)$ for every *invariant* polynomial $P$?

Algebra

Is $0 \in \overline{Gv}$?

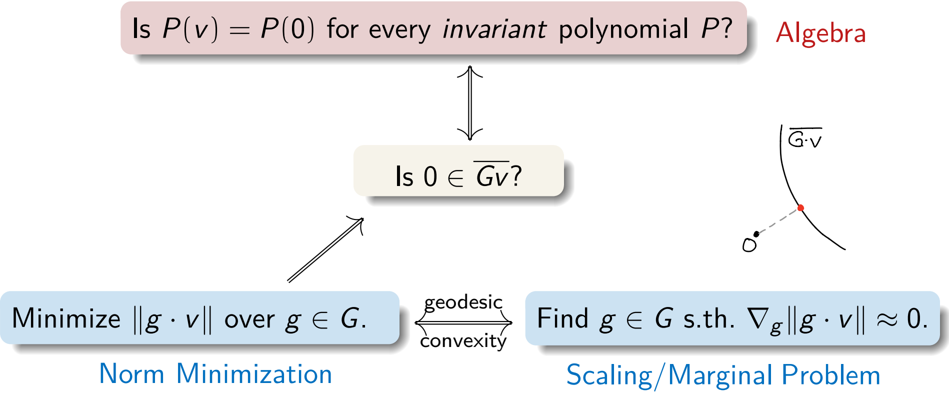
Minimize $\|g \cdot v\|$ over $g \in G$.

Norm Minimization

Find $g \in G$ s.th. $\nabla_g \|g \cdot v\| \approx 0$.

Scaling/Marginal Problem

# Big picture: Null cone, optimization, and scaling



Is $P(v) = P(0)$ for every *invariant* polynomial $P$? Algebra

$\Updownarrow$

Is $0 \in \overline{Gv}$?

$\Uparrow$

Minimize $\|g \cdot v\|$ over $g \in G$. $\xLeftrightarrow[\text{convexity}]{\text{geodesic}}$ Find $g \in G$ s.th. $\nabla_g \|g \cdot v\| \approx 0$.

Norm Minimization

Scaling/Marginal Problem

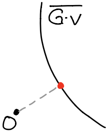
# Big picture: Null cone, optimization, and scaling

Is $P(v) = P(0)$ for every *invariant* polynomial $P$?

Algebra

$\Updownarrow$

Is $0 \in \overline{Gv}$?



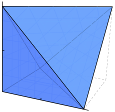
⇗

Minimize $\|g \cdot v\|$ over $g \in G$. $\xLeftrightarrow[\text{convexity}]{\text{geodesic}}$ Find $g \in G$ s.th. $\nabla_g \|g \cdot v\| \approx 0$.

Norm Minimization

Scaling/Marginal Problem

$\Updownarrow$



Polytopes

Why care? Applications, plausibly solvable in poly time, offers path to other orbit problems. . . **let's get started!**

# Hold on, isn't this is a Quantum Information workshop!?

Two motivations:

1. Quantum problems: Orbit problems (for noncommutative groups) capture interesting quantum information applications.
2. Quantum solutions: Quantum computers are good at linear algebra and optimization. Can they solve orbit problems faster?

Today we focus on the first direction. Harold will discuss the second tomorrow.

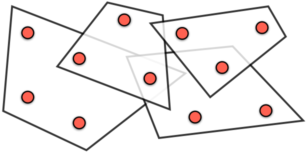
# Hold on, isn't this is a Quantum Information workshop!?

Two motivations:

1. Quantum problems: Orbit problems (for noncommutative groups) capture interesting quantum information applications.
2. Quantum solutions: Quantum computers are good at linear algebra and optimization. Can they solve orbit problems faster?

Today we focus on the first direction. Harold will discuss the second tomorrow.

# A panorama of applications



# Example: Matrix scaling revisited

Let $G = \mathrm{T}_n \times \mathrm{T}_n$ act on $V = \mathrm{Mat}_n(\mathbb{C})$ by row-column scaling:

$$(g, h) \cdot M = \begin{pmatrix} g_1 & & \\ & \ddots & \\ & & g_n \end{pmatrix} M \begin{pmatrix} h_1 & & \\ & \ddots & \\ & & h_n \end{pmatrix}$$

*Norm minimization:*

$$\inf_{g,h} \|(g, h) \cdot M\|^2 = \inf_{g,h} \sum_{i,j} |g_i M_{ij} h_j|^2 = \inf_{x,y \in \mathbb{R}^n} \sum_{i,j} |M_{ij}|^2 e^{x_i + y_j}$$

- geometric program, log-convex in $x$, $y$

*Gradient*:

$$\nabla_{x=y=0} \log(\ldots) = \frac{1}{\|M\|^2} (\mathbf{r}(M), \mathbf{c}(M))$$

where $\mathbf{r}(M)$, $\mathbf{c}(M)$ row and column sums of matrix with entries $|M_{ij}|^2$.

Norm minimization and matrix scaling are equivalent by convexity! ☺
Motivates why Sinkhorn works and starting point for cutting-edge algos.

# Example: Matrix scaling revisited

Let $G = \mathrm{T}_n \times \mathrm{T}_n$ act on $V = \mathrm{Mat}_n(\mathbb{C})$ by row-column scaling:

$$(g,h) \cdot M = \begin{pmatrix} g_1 & & \\ & \ddots & \\ & & g_n \end{pmatrix} M \begin{pmatrix} h_1 & & \\ & \ddots & \\ & & h_n \end{pmatrix}$$

*Norm minimization:*

$$\inf_{g,h} \|(g,h) \cdot M\|^2 = \inf_{g,h} \sum_{i,j} |g_i M_{ij} h_j|^2 = \inf_{x,y \in \mathbb{R}^n} \sum_{i,j} |M_{ij}|^2 e^{x_i + y_j}$$

- geometric program, log-convex in $x$, $y$

*Gradient*:

$$\nabla_{x=y=0} \log(\dots) = \frac{1}{\|M\|^2}(\mathbf{r}(M), \mathbf{c}(M))$$

where $\mathbf{r}(M)$, $\mathbf{c}(M)$ row and column sums of matrix with entries $|M_{ij}|^2$.

Norm minimization and matrix scaling are equivalent by convexity! ☺
Motivates why Sinkhorn works and starting point for cutting-edge algos.

# Example: Matrix scaling revisited

Let $G = \mathrm{T}_n \times \mathrm{T}_n$ act on $V = \mathrm{Mat}_n(\mathbb{C})$ by row-column scaling:

$$(g, h) \cdot M = \begin{pmatrix} g_1 & & \\ & \ddots & \\ & & g_n \end{pmatrix} M \begin{pmatrix} h_1 & & \\ & \ddots & \\ & & h_n \end{pmatrix}$$

*Norm minimization:*

$$\inf_{g,h} \|(g, h) \cdot M\|^2 = \inf_{g,h} \sum_{i,j} |g_i M_{ij} h_j|^2 = \inf_{x,y \in \mathbb{R}^n} \sum_{i,j} |M_{ij}|^2 e^{x_i + y_j}$$

- geometric program, log-convex in $x$, $y$

*Gradient*:

$$\nabla_{x=y=0} \log(\dots) = \frac{1}{\|M\|^2} (\mathbf{r}(M), \mathbf{c}(M))$$

where $\mathbf{r}(M)$, $\mathbf{c}(M)$ row and column sums of matrix with entries $|M_{ij}|^2$.

Norm minimization and matrix scaling are equivalent by convexity! ☺
Motivates why Sinkhorn works and starting point for cutting-edge algos.

# Example: Matrix scaling revisited

Let $G = \mathrm{T}_n \times \mathrm{T}_n$ act on $V = \mathrm{Mat}_n(\mathbb{C})$ by row-column scaling:

$$(g,h) \cdot M = \begin{pmatrix} g_1 & & \\ & \ddots & \\ & & g_n \end{pmatrix} M \begin{pmatrix} h_1 & & \\ & \ddots & \\ & & h_n \end{pmatrix}$$

*Norm minimization:*

$$\inf_{g,h} \|(g,h) \cdot M\|^2 = \inf_{g,h} \sum_{i,j} |g_i M_{ij} h_j|^2 = \inf_{x,y \in \mathbb{R}^n} \sum_{i,j} |M_{ij}|^2 e^{x_i + y_j}$$

- geometric program, log-convex in $x$, $y$

*Gradient*:

$$\nabla_{x=y=0} \log(\dots) = \frac{1}{\|M\|^2} \big(\mathbf{r}(M), \mathbf{c}(M)\big)$$

where $\mathbf{r}(M)$, $\mathbf{c}(M)$ row and column sums of matrix with entries $|M_{ij}|^2$.

Norm minimization and matrix scaling are equivalent by convexity! ☺
Motivates why Sinkhorn works and starting point for cutting-edge algos.

# Example: Matrix scaling revisited

Let $G = \mathrm{ST}_n \times \mathrm{ST}_n$ act on $V = \mathrm{Mat}_n(\mathbb{C})$ by row-column scaling:

$$(g, h) \cdot M = \begin{pmatrix} g_1 & & \\ & \ddots & \\ & & g_n \end{pmatrix} M \begin{pmatrix} h_1 & & \\ & \ddots & \\ & & h_n \end{pmatrix}$$

*Norm minimization:*

$$\inf_{g,h} \|(g, h) \cdot M\|^2 = \inf_{g,h} \sum_{i,j} |g_i M_{ij} h_j|^2 = \inf_{x,y \in \mathbb{R}^n_{\sum=0}} \sum_{i,j} |M_{ij}|^2 e^{x_i + y_j}$$

- geometric program, log-convex in $x$, $y$

*Gradient*:

$$\nabla_{x=y=0} \log(\dots) = \frac{1}{\|M\|^2}\big(\mathbf{r}(M), \mathbf{c}(M)\big) - \frac{1}{n}(\mathbf{1}, \mathbf{1})$$

where $\mathbf{r}(M)$, $\mathbf{c}(M)$ row and column sums of matrix with entries $|M_{ij}|^2$.

Norm minimization and matrix scaling are equivalent by convexity! ☺
Motivates why Sinkhorn works and starting point for cutting-edge algos.

# Example: Laurent polynomials

Let $G = \mathrm{T}_n$ act on Laurent polynomials in $n$ variables by scaling variables:

$$P = \sum_{\omega \in \mathbb{Z}^n} p_\omega Z^\omega \qquad \Rightarrow \qquad g \cdot P = \sum_{\omega \in \mathbb{Z}^n} p_\omega g^\omega Z^\omega$$

*Norm minimization:*

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

- again geometric program, log-convex in $x$
- $\inf > 0$ iff $0 \in$ Newton polytope of $P$

*Gradient:*

$$\nabla_{x=0} \log(\dots) = \frac{\sum_\omega |p_\omega|^2 \, \omega}{\sum_\omega |p_\omega|^2} \in \Delta(P)$$

Captures linear and (unconstrained) geometric programming!

# Example: Laurent polynomials

Let $G = \mathrm{T}_n$ act on Laurent polynomials in $n$ variables by scaling variables:

$$P = \sum_{\omega \in \mathbb{Z}^n} p_\omega Z^\omega \qquad \Rightarrow \qquad g \cdot P = \sum_{\omega \in \mathbb{Z}^n} p_\omega g^\omega Z^\omega$$

*Norm minimization:*

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

- again geometric program, log-convex in $x$
- $\inf > 0$ iff $0 \in$ Newton polytope of $P$

*Gradient:*

$$\nabla_{x=0} \log(\dots) = \frac{\sum_\omega |p_\omega|^2 \, \omega}{\sum_\omega |p_\omega|^2} \in \Delta(P)$$

Captures linear and (unconstrained) geometric programming!

# Example: Laurent polynomials

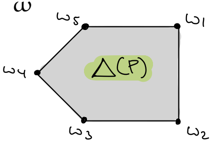
Let $G = \mathrm{T}_n$ act on Laurent polynomials in $n$ variables by scaling variables:

$$P = \sum_{\omega \in \mathbb{Z}^n} p_\omega Z^\omega \qquad \Rightarrow \qquad g \cdot P = \sum_{\omega \in \mathbb{Z}^n} p_\omega g^\omega Z^\omega$$

*Norm minimization:*

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

- again geometric program, log-convex in $x$
- $\inf > 0$ iff $0 \in$ Newton polytope of $P$



*Gradient:* $$\nabla_{x=0} \log(\dots) = \frac{\sum_\omega |p_\omega|^2 \, \omega}{\sum_\omega |p_\omega|^2} \in \Delta(P)$$

Captures linear and (unconstrained) geometric programming!

# Example: Laurent polynomials

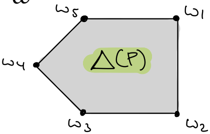
Let $G = \mathrm{T}_n$ act on Laurent polynomials in $n$ variables by scaling variables:

$$P = \sum_{\omega \in \mathbb{Z}^n} p_\omega Z^\omega \qquad \Rightarrow \qquad g \cdot P = \sum_{\omega \in \mathbb{Z}^n} p_\omega g^\omega Z^\omega$$

*Norm minimization:*

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

- again geometric program, log-convex in $x$
- inf $> 0$ iff $0 \in$ Newton polytope of $P$



*Gradient:*

$$\boxed{\nabla_{x=0} \log(\dots) = \frac{\sum_\omega |p_\omega|^2 \, \omega}{\sum_\omega |p_\omega|^2}} \in \Delta(P)$$

Captures linear and (unconstrained) geometric programming!

# Example: Operator and tensor scaling

What might a quantum version of the matrix scaling problem look like?

For an operator $\rho \in \mathrm{PSD}(\mathbb{C}^n \otimes \mathbb{C}^n)$, say a *scaling* is of the form

$$\sigma = (g \otimes h)\rho(g^* \otimes h^*) \qquad (g, h \in \mathrm{GL}_n).$$

Operator scaling problem: Given $\rho$, find scaling such that $\sigma_1, \sigma_2 \approx I$.

Interesting to generalize to more tensor factors and arbitrary marginals:

Tensor scaling (quant. marginal) problem: Given $\rho$, which $(\sigma_1, \ldots, \sigma_d)$ can be obtained by scaling?

- eigenvalues form *convex polytopes* with exp. many vertices and facets
- encode local info about entanglement, tensor ranks, ...

Key challenge: Can we find efficient *algorithmic* solution?

# Example: Operator and tensor scaling

What might a quantum version of the matrix scaling problem look like? For an operator $\rho \in \mathrm{PSD}(\mathbb{C}^n \otimes \mathbb{C}^n)$, say a *scaling* is of the form

$$\sigma = (g \otimes h)\rho(g^* \otimes h^*) \qquad (g, h \in \mathrm{GL}_n).$$

Operator scaling problem: Given $\rho$, find scaling such that $\sigma_1, \sigma_2 \approx I$.

Interesting to generalize to more tensor factors and arbitrary marginals:

Tensor scaling (quant. marginal) problem: Given $\rho$, which $(\sigma_1, \dots, \sigma_d)$ can be obtained by scaling?

- eigenvalues form *convex polytopes* with exp. many vertices and facets
- encode local info about entanglement, tensor ranks, ...

Key challenge: Can we find efficient *algorithmic* solution?

# Example: Operator and tensor scaling

What might a quantum version of the matrix scaling problem look like? For an operator $\rho \in \mathrm{PSD}(\mathbb{C}^n \otimes \mathbb{C}^n)$, say a *scaling* is of the form

$$\sigma = (g \otimes h)\rho(g^* \otimes h^*) \qquad (g, h \in \mathrm{GL}_n).$$

Operator scaling problem: Given $\rho$, find scaling such that $\sigma_1, \sigma_2 \approx I$.

Interesting to generalize to more tensor factors and arbitrary marginals:

Tensor scaling (quant. marginal) problem: Given $\rho$, which $(\sigma_1, \ldots, \sigma_d)$ can be obtained by scaling?

- eigenvalues form *convex polytopes* with exp. many vertices and facets
- encode local info about entanglement, tensor ranks, . . .

Key challenge: Can we find efficient *algorithmic* solution?

# Example: Operator and tensor scaling

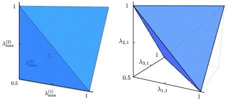
What might a quantum version of the matrix scaling problem look like? For an operator $\rho \in \mathrm{PSD}(\mathbb{C}^n \otimes \mathbb{C}^n)$, say a *scaling* is of the form

$$\sigma = (g \otimes h)\rho(g^* \otimes h^*) \qquad (g, h \in \mathrm{GL}_n).$$

Operator scaling problem: Given $\rho$, find scaling such that $\sigma_1, \sigma_2 \approx I$.

Interesting to generalize to more tensor factors and arbitrary marginals:

Tensor scaling (quant. marginal) problem: Given $\rho$, which $(\sigma_1, \dots, \sigma_d)$ can be obtained by scaling?



- eigenvalues form *convex polytopes* with exp. many vertices and facets
- encode local info about entanglement, tensor ranks, . . .

Key challenge: Can we find efficient *algorithmic* solution?

# Example: Operator scaling and polynomial identity testing

We can interpret $\rho, \sigma$ as Choi operators of completely positive maps

$$\Phi(A) = \sum_k X_k A X_k{}^*, \qquad \Psi(A) = \sum_k Y_k A Y_k{}^*.$$

*Scaling* translates into left-right action on Kraus operators: $Y_k = g X_k h^T$.
The condition $\sigma_1 = \sigma_2 = I$ means that $\Psi$ is unital and trace-preserving.

Operator scaling problem: Given $\Phi$, find unital & trace-preserving scaling.

Algebra: Possible iff $\exists$ matrices $c_k$ s.th. $\det \sum_k c_k \otimes X_k \neq 0$.

- matrix $X(c) = \sum_k c_k X_k$ in NC variables $c_k$ has *maximal NC-rank*
- when $c_k$ restricted to scalars: *major open problem in TCS!*

Operator scaling can be solved in deterministic poly-time [Garg et al, cf. Ivanyos et al]

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

# Example: Operator scaling and polynomial identity testing

We can interpret $\rho, \sigma$ as Choi operators of completely positive maps

$$\Phi(A) = \sum_k X_k A X_k{}^*, \qquad \Psi(A) = \sum_k Y_k A Y_k{}^*.$$

*Scaling* translates into left-right action on Kraus operators: $Y_k = g X_k h^T$.
The condition $\sigma_1 = \sigma_2 = I$ means that $\Psi$ is unital and trace-preserving.

Operator scaling problem: Given $\Phi$, find unital & trace-preserving scaling.

Algebra: Possible iff $\exists$ matrices $c_k$ s.th. $\det \sum_k c_k \otimes X_k \neq 0$.

- matrix $X(c) = \sum_k c_k X_k$ in NC variables $c_k$ has *maximal NC-rank*
- when $c_k$ restricted to scalars: *major open problem in TCS!*

Operator scaling can be solved in deterministic poly-time [Garg et al, cf. Ivanyos et al]

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

# Example: Operator scaling and polynomial identity testing

We can interpret $\rho, \sigma$ as Choi operators of completely positive maps

$$\Phi(A) = \sum_k X_k A X_k{}^*, \qquad \Psi(A) = \sum_k Y_k A Y_k{}^*.$$

*Scaling* translates into left-right action on Kraus operators: $Y_k = g X_k h^T$.
The condition $\sigma_1 = \sigma_2 = I$ means that $\Psi$ is unital and trace-preserving.

Operator scaling problem: Given $\Phi$, find unital & trace-preserving scaling.

Algebra: Possible iff $\exists$ matrices $c_k$ s.th. $\det \sum_k c_k \otimes X_k \neq 0$.

- matrix $X(c) = \sum_k c_k X_k$ in NC variables $c_k$ has *maximal NC-rank*
- when $c_k$ restricted to scalars: *major open problem in TCS!*

Operator scaling can be solved in deterministic poly-time [Garg et al, cf. Ivanyos et al]

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

# Example: Operator scaling and polynomial identity testing

We can interpret $\rho, \sigma$ as Choi operators of completely positive maps

$$\Phi(A) = \sum_k X_k A X_k^*, \qquad \Psi(A) = \sum_k Y_k A Y_k^*.$$

*Scaling* translates into left-right action on Kraus operators: $Y_k = g X_k h^T$.
The condition $\sigma_1 = \sigma_2 = I$ means that $\Psi$ is unital and trace-preserving.

Operator scaling problem: Given $\Phi$, find unital & trace-preserving scaling.

Algebra: Possible iff $\exists$ matrices $c_k$ s.th. $\det \sum_k c_k \otimes X_k \neq 0$.

- matrix $X(c) = \sum_k c_k X_k$ in NC variables $c_k$ has *maximal NC-rank*
- when $c_k$ restricted to scalars: *major open problem in TCS!*

Operator scaling can be solved in deterministic poly-time [Garg et al, cf. Ivanyos et al]

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

# Connections and applications

Invariant theory: Null cone & orbit closure intersection, moment polytopes

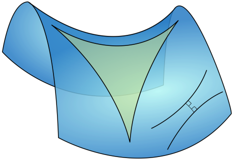
Analysis: Brascamp-Lieb inequalities, solution of Paulsen's problem

Symplectic geometry: Horn's problem $\exists\ A + B = C$ with spectrum $\alpha$, $\beta$, $\gamma$?
Combinatorics: Positivity of Littlewood-Richardson coefficients

Statistics: MLE in Gaussian models, Tyler M-estimator
Machine Learning: Optimal transport
Optimization: Efficient algorithms for class of quadratic equations

Computational complexity: Polynomial identity testing, tensor ranks
Quantum information: Marginal problems, entanglement transformations, normal forms of tensors

# From Euclidean to geodesic convexity



# Norm minimization and gradient

We want to minimize the function:

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

Consider $G = \mathrm{GL}_n$. By the polar decomposition, if $U_n$ preserves the norm we can restrict the minimization to:

$$\mathrm{PD}_n = \{p = e^X \ : \ X \in \mathrm{Herm}_n\}$$

We can use this change of variables to define gradient at $p = I$:

$$\nabla_{X=0} F(e^X)$$

- arises from natural Riemannian metric on $\mathrm{PD}_n$
- known as *moment map* in Hamiltonian physics, symplectic geometry
- It turns out that $\mu(v) = 0$ captures natural scaling problems!

Analogously for, e.g., $G = \mathrm{SL}_n \rightsquigarrow X$ traceless.

# Norm minimization and gradient

We want to minimize the function:

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

Consider $G = \mathrm{GL}_n$. By the polar decomposition, if $U_n$ preserves the norm we can restrict the minimization to:

$$\mathrm{PD}_n = \{p = e^X \ : \ X \in \mathrm{Herm}_n\}$$

We can use this change of variables to define gradient at $p = I$:

$$\nabla_{X=0} F(e^X)$$

- arises from natural Riemannian metric on $\mathrm{PD}_n$
- known as *moment map* in Hamiltonian physics, symplectic geometry
- It turns out that $\mu(v) = 0$ captures natural scaling problems!

Analogously for, e.g., $G = \mathrm{SL}_n \rightsquigarrow X$ traceless.

# Norm minimization and gradient

We want to minimize the function:

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

Consider $G = \mathrm{GL}_n$. By the polar decomposition, if $U_n$ preserves the norm we can restrict the minimization to:

$$\mathrm{PD}_n = \{p = e^X \ : \ X \in \mathrm{Herm}_n\}$$

We can use this change of variables to define gradient at $p = I$:

$$\nabla_{X=0} F(e^X)$$

- arises from natural Riemannian metric on $\mathrm{PD}_n$
- known as *moment map* in Hamiltonian physics, symplectic geometry
- It turns out that $\mu(v) = 0$ captures natural scaling problems!

Analogously for, e.g., $G = \mathrm{SL}_n \rightsquigarrow X$ traceless.

# Norm minimization and gradient

We want to minimize the function:

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

Consider $G = \mathrm{GL}_n$. By the polar decomposition, if $U_n$ preserves the norm we can restrict the minimization to:

$$\mathrm{PD}_n = \{p = e^X \ : \ X \in \mathrm{Herm}_n\}$$

We can use this change of variables to define gradient at $p = I$:

$$\mu\colon V \setminus \{0\} \to \mathrm{Herm}_n, \quad \mu(v) = \nabla_{X=0} F(e^X)$$

- arises from natural Riemannian metric on $\mathrm{PD}_n$
- known as *moment map* in Hamiltonian physics, symplectic geometry
- It turns out that $\mu(v) = 0$ captures natural scaling problems!

Analogously for, e.g., $G = \mathrm{SL}_n \rightsquigarrow X$ traceless.

# Norm minimization and gradient

We want to minimize the function:

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

Consider $G = \mathrm{GL}_n$. By the polar decomposition, if $U_n$ preserves the norm we can restrict the minimization to:

$$\mathrm{PD}_n = \{p = e^X \ : \ X \in \mathrm{Herm}_n\}$$

We can use this change of variables to define gradient at $p = I$:

$$\mu\colon V \setminus \{0\} \to \mathrm{Herm}_n, \quad \mu(v) = \nabla_{X=0} F(e^X)$$

- arises from natural Riemannian metric on $\mathrm{PD}_n$
- known as *moment map* in Hamiltonian physics, symplectic geometry
- It turns out that $\mu(v) = 0$ captures natural scaling problems!

Analogously for, e.g., $G = \mathrm{SL}_n \rightsquigarrow X$ traceless.

# Example: Operator scaling revisited

Let $G = \mathrm{GL}_n \times \mathrm{GL}_n$ act on $V = \mathrm{Lin}(\mathbb{C}^n \otimes \mathbb{C}^n)$:

$$(g, h) \cdot M = (g \otimes h)M$$

*Norm minimization:*

$$\inf_{g,h} \|(g \otimes h)M\|_F^2 = \inf_{X,Y \in \mathrm{Herm}_n} \mathrm{tr}(e^X \otimes e^Y)MM^*$$

*Gradient:*

$$\mu(M) = \nabla_{X=Y=0} \log(\dots) = \frac{1}{\mathrm{tr}\,\rho}(\rho_1, \rho_2)$$

where $\rho = MM^*$. Restricted to $G = \mathrm{SL}_n \times \mathrm{SL}_n$: captures operator scaling!

However, the objective is *not* convex in $X$, $Y$...

# Example: Operator scaling revisited

Let $G = \mathrm{GL}_n \times \mathrm{GL}_n$ act on $V = \mathrm{Lin}(\mathbb{C}^n \otimes \mathbb{C}^n)$:

$$(g, h) \cdot M = (g \otimes h) M$$

*Norm minimization:*

$$\inf_{g,h} \|(g \otimes h) M\|_F^2 = \inf_{X,Y \in \mathrm{Herm}_n} \mathrm{tr}(e^X \otimes e^Y) M M^*$$

*Gradient:*

$$\mu(M) = \nabla_{X=Y=0} \log(\dots) = \frac{1}{\mathrm{tr}\,\rho}(\rho_1, \rho_2)$$

where $\rho = MM^*$. Restricted to $G = \mathrm{SL}_n \times \mathrm{SL}_n$: captures operator scaling!

However, the objective is *not* convex in $X$, $Y$...

# Example: Operator scaling revisited

Let $G = \mathrm{GL}_n \times \mathrm{GL}_n$ act on $V = \mathrm{Lin}(\mathbb{C}^n \otimes \mathbb{C}^n)$:

$$(g, h) \cdot M = (g \otimes h) M$$

*Norm minimization:*

$$\inf_{g,h} \|(g \otimes h) M\|_F^2 = \inf_{X,Y \in \mathrm{Herm}_n} \mathrm{tr}(e^X \otimes e^Y) M M^*$$

*Gradient:*

$$\mu(M) = \nabla_{X=Y=0} \log(\ldots) = \frac{1}{\mathrm{tr}\,\rho} (\rho_1, \rho_2)$$

where $\rho = MM^*$. Restricted to $G = \mathrm{SL}_n \times \mathrm{SL}_n$: captures operator scaling!

However, the objective is *not* convex in $X$, $Y$...

# Example: Operator scaling revisited

Let $G = \mathrm{GL}_n \times \mathrm{GL}_n$ act on $V = \mathrm{Lin}(\mathbb{C}^n \otimes \mathbb{C}^n)$:

$$(g, h) \cdot M = (g \otimes h) M$$

*Norm minimization:*

$$\inf_{g,h} \|(g \otimes h) M\|_F^2 = \inf_{X,Y \in \mathrm{Herm}_n} \operatorname{tr}(e^X \otimes e^Y) M M^*$$

*Gradient:*

$$\mu(M) = \nabla_{X=Y=0} \log(\ldots) = \frac{1}{\operatorname{tr} \rho} (\rho_1, \rho_2)$$

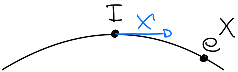
where $\rho = MM^*$. Restricted to $G = \mathrm{SL}_n \times \mathrm{SL}_n$: captures operator scaling!

However, the objective is *not* convex in $X$, $Y$...

# Geodesic convexity

$$F\colon \mathrm{PD}_n \to \mathbb{R}, \quad F(p) := \log \|p \cdot v\|$$

is convex along the curves $e^{Xt}$ for $X \in \mathrm{Herm}_n$, which are geodesics of $\mathrm{PD}_n$. That is, $F$ is geodesically convex!



*Proof?* $\{e^{Xt}\} \Rightarrow$ *commutative subgroup* $\Rightarrow$ *Laurent polynomials* ☺

Just like in the Euclidean case, convexity implies critical points are minima:

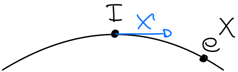
$$\|v\| = \inf_g \|g \cdot v\| \qquad \Leftrightarrow \qquad \mu(v) \equiv \nabla_{X=0} F(e^X) = 0$$

How convex for given action? Necessary for algorithms!

# Geodesic convexity

$$F \colon \mathrm{PD}_n \to \mathbb{R}, \quad F(p) := \log \|p \cdot v\|$$

is convex along the curves $e^{Xt}$ for $X \in \mathrm{Herm}_n$, which are geodesics of $\mathrm{PD}_n$. That is, $F$ is geodesically convex!



*Proof?* $\{e^{Xt}\} \Rightarrow$ *commutative subgroup* $\Rightarrow$ *Laurent polynomials* ☺

Just like in the Euclidean case, convexity implies critical points are minima:

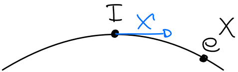
$$\|v\| = \inf_g \|g \cdot v\| \qquad \Leftrightarrow \qquad \mu(v) \equiv \nabla_{X=0} F(e^X) = 0$$

How convex for given action? Necessary for algorithms!

# Geodesic convexity

$$F\colon \mathrm{PD}_n \to \mathbb{R}, \quad F(p) := \log \|p \cdot v\|$$

is convex along the curves $e^{Xt}$ for $X \in \mathrm{Herm}_n$, which are geodesics of $\mathrm{PD}_n$. That is, $F$ is geodesically convex!



*Proof?* $\{e^{Xt}\} \Rightarrow$ *commutative subgroup* $\Rightarrow$ *Laurent polynomials* ☺

Just like in the Euclidean case, convexity implies critical points are minima:

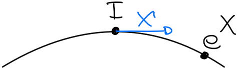
$$\|v\| = \inf_g \|g \cdot v\| \qquad \Leftrightarrow \qquad \mu(v) \equiv \nabla_{X=0} F(e^X) = 0$$

How convex for given action? Necessary for algorithms!

# Geodesic convexity

$$F\colon \mathrm{PD}_n \to \mathbb{R}, \quad F(p) := \log \|p \cdot v\|$$

is convex along the curves $e^{Xt}$ for $X \in \mathrm{Herm}_n$, which are geodesics of $\mathrm{PD}_n$. That is, $F$ is geodesically convex!



*Proof?* $\{e^{Xt}\} \Rightarrow$ *commutative subgroup* $\Rightarrow$ *Laurent polynomials* ☺

Just like in the Euclidean case, convexity implies critical points are minima:

$$\|v\| = \inf_g \|g \cdot v\| \qquad \Leftrightarrow \qquad \mu(v) \equiv \nabla_{X=0} F(e^X) = 0$$

How convex for given action? Necessary for algorithms!

# Geodesic convexity made quantitative

The objective $F(p) = \log \|p \cdot v\|$ is geodesically smooth, meaning

$$\partial_t^2 F(e^{Xt}) \leqslant L \|X\|_F^2.$$

**Theorem: Noncommutative duality estimates**

$$1 - \frac{\|\mu(v)\|_F}{\gamma} \leqslant \frac{\inf_g \|g \cdot v\|^2}{\|v\|^2} \leqslant 1 - \frac{\|\mu(v)\|_F^2}{2L}$$

- ☺ norm minimization ⇔ scaling in a quantitative way
- ☺ null cone membership reduces to solving either
- ☺ scaling is possible *iff* not in null cone [Kempf-Ness '79]

Parameters $L$, $\gamma$ depend on combinatorial data of action.

# Geodesic convexity made quantitative

The objective $F(p) = \log \|p \cdot v\|$ is geodesically smooth, meaning

$$\partial_t^2 F(e^{Xt}) \leqslant L\|X\|_F^2.$$

**Theorem: Noncommutative duality estimates**

$$1 - \frac{\|\mu(v)\|_F}{\gamma} \leqslant \frac{\inf_g \|g \cdot v\|^2}{\|v\|^2} \leqslant 1 - \frac{\|\mu(v)\|_F^2}{2L}$$

- ☺ norm minimization ⇔ scaling in a quantitative way
- ☺ null cone membership reduces to solving either
- ☺ scaling is possible *iff* not in null cone [Kempf-Ness '79]

Parameters $L$, $\gamma$ depend on combinatorial data of action.

# Noncommutative duality

The objective $F(p) = \log \|p \cdot v\|$ is geodesically smooth, meaning

$$\partial_t^2 F(e^{Xt}) \leqslant L \|X\|_F^2.$$

**Theorem: Noncommutative duality estimates**

$$1 - \frac{\|\mu(v)\|_F}{\gamma} \leqslant \frac{\inf_g \|g \cdot v\|^2}{\|v\|^2} \leqslant 1 - \frac{\|\mu(v)\|_F^2}{2L}$$

- ☺ norm minimization ⇔ scaling in a quantitative way
- ☺ null cone membership reduces to solving either
- ☺ scaling is possible *iff* not in null cone [Kempf-Ness '79]

Parameters $L$, $\gamma$ depend on combinatorial data of action.

# Noncommutative duality

The objective $F(p) = \log \|p \cdot v\|$ is geodesically smooth, meaning

$$\partial_t^2 F(e^{Xt}) \leqslant L \|X\|_F^2.$$

**Theorem: Noncommutative duality estimates**

$$1 - \frac{\|\mu(v)\|_F}{\gamma} \leqslant \frac{\inf_g \|g \cdot v\|^2}{\|v\|^2} \leqslant 1 - \frac{\|\mu(v)\|_F^2}{2L}$$

- ☺ norm minimization ⇔ scaling in a quantitative way
- ☺ null cone membership reduces to solving either
- ☺ scaling is possible *iff* not in null cone [Kempf-Ness '79]

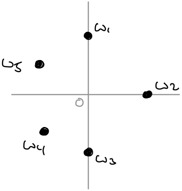
Parameters $L$, $\gamma$ depend on combinatorial data of action.

# Interlude: Weights of action

Take any action of $\mathrm{GL}_n$. If we restrict to $\mathrm{T}_n = \left(\begin{smallmatrix} \cdot & \\ & \cdot_{\cdot} \end{smallmatrix}\right)$, can find basis of $V \cong \mathbb{C}^m$ s.th. action equivalent to scaling Laurent polys. The exponents

$$\Omega = \{\omega_1, \ldots, \omega_m\} \subseteq \mathbb{Z}^n.$$

are called weights, and they completely characterize the action.



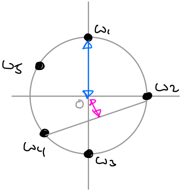
Their geometry determine the geodesic convexity parameters $L$, $\gamma$.

# Interlude: Weights of action

Take any action of $\mathrm{GL}_n$. If we restrict to $\mathrm{T}_n = \left( \ddots \right)$, can find basis of $V \cong \mathbb{C}^m$ s.th. action equivalent to scaling Laurent polys. The exponents

$$\Omega = \{\omega_1, \ldots, \omega_m\} \subseteq \mathbb{Z}^n.$$

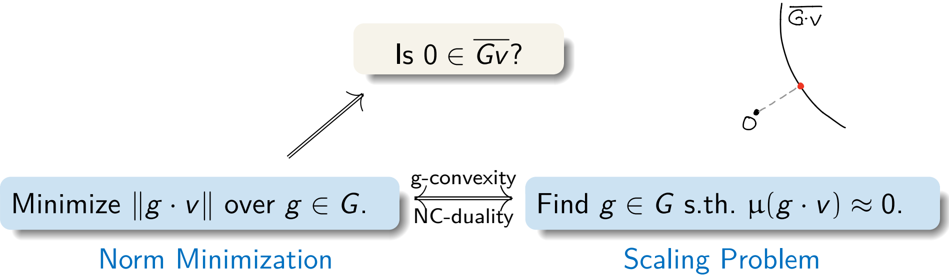
are called weights, and they completely characterize the action.



Their geometry determine the geodesic convexity parameters $L$, $\gamma$.

# Summary so far: Noncommutative group optimization [BFGOWW]

Action of “nice” $G \subseteq \mathrm{GL}_n$ on $V \cong \mathbb{C}^m$, $\quad \mu(v) = \nabla_{p=I} \log \|p \cdot v\|$.



Is $0 \in \overline{Gv}$?

Minimize $\|g \cdot v\|$ over $g \in G$. $\xLeftrightarrow[\text{NC-duality}]{\text{g-convexity}}$ Find $g \in G$ s.th. $\mu(g \cdot v) \approx 0$.

Norm Minimization

Scaling Problem

- Geodesic convexity explains why simple greedy algorithms can work.
- Scaling, norm minimization, and null cone related in *quantitative* way.
- Non-commutative generalization of convex programming duality.
- All examples mentioned before fall into this framework.

# Interlude: Beyond $\mathsf{GL}_n$ and $\mathsf{SL}_n$

All the preceding generalizes to complex reductive groups – not just $\mathsf{SL}_n$, $\mathsf{T}_n$, $\mathsf{ST}_n$, and products thereof. Concretely, this means a subgroup

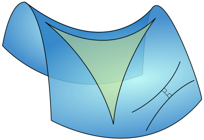
$$G \subseteq \mathsf{GL}_n(\mathbb{C})$$

defined by polynomial equations that is closed under taking adjoints.

Any such group has a polar decomposition $g = up$, so we can reduce to

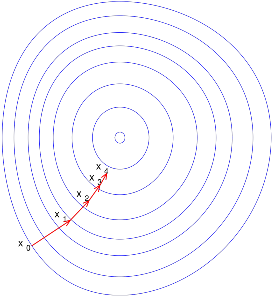
$$G \cap \mathsf{PD}_n = \{g^* g : g \in G\}.$$

This is a Hadamard manifold (in fact a symmetric space of noncompact type), a particularly nice Riemannian manifold of nonpositive curvature.



However, nonpositive curvature poses unique challenges for optimization.

# Algorithms



# First order algorithm for scaling (“gradient descent”)

Idea: Repeatedly perform geodesic gradient steps

$$g \leftarrow e^{-\frac{1}{L}\nabla F(g)} g = e^{-\frac{1}{L}\mu(g\cdot v)} g.$$

**Theorem**

Let $v \in V$ be not in the null cone. Then the algorithm outputs $g \in G$ such that $\|\mu(g \cdot v)\| \leqslant \varepsilon$ within $T = \text{poly}(\frac{1}{\varepsilon}, \text{input size})$ steps.

Analysis: Smoothness implies $F$ decreases in each step. Combine with a priori lower bound obtained using constructive invariant theory.

**Corollary**

Same algorithm solves null cone problem in time $\text{poly}(\frac{1}{\gamma}, \text{input size})$.

# First order algorithm for scaling (“gradient descent”)

Idea: Repeatedly perform geodesic gradient steps

$$g \leftarrow e^{-\frac{1}{L}\nabla F(g)} g = e^{-\frac{1}{L}\mu(g\cdot v)} g.$$

**Theorem**

Let $v \in V$ be not in the null cone. Then the algorithm outputs $g \in G$ such that $\|\mu(g \cdot v)\| \leqslant \varepsilon$ within $T = \text{poly}(\frac{1}{\varepsilon}, \text{input size})$ steps.

Analysis: Smoothness implies $F$ decreases in each step. Combine with a priori lower bound obtained using constructive invariant theory.

**Corollary**

Same algorithm solves null cone problem in time $\text{poly}(\frac{1}{\gamma}, \text{input size})$.

# First order algorithm for scaling ("gradient descent")

Idea: Repeatedly perform geodesic gradient steps

$$g \leftarrow e^{-\frac{1}{L}\nabla F(g)} g = e^{-\frac{1}{L}\mu(g\cdot v)} g.$$

**Theorem**

Let $v \in V$ be not in the null cone. Then the algorithm outputs $g \in G$ such that $\|\mu(g \cdot v)\| \leqslant \varepsilon$ within $T = \text{poly}(\frac{1}{\varepsilon}, \text{input size})$ steps.

Analysis: Smoothness implies $F$ decreases in each step. Combine with a priori lower bound obtained using constructive invariant theory.

**Corollary**

Same algorithm solves null cone problem in time $\text{poly}(\frac{1}{\gamma}, \text{input size})$.

# Second order algorithm for norm minimization

Rough idea: Minimize local quadratic approximation (after regularization)

$$Q(H) = F(g) + \nabla F(g)[H] + \frac{1}{2}\nabla^2 F(g)[H, H] \quad \approx \quad F(e^H g)$$

on small neighborhoods, where it can be trusted. Need $F$ "robust".

**Theorem**

Let $v \in V$ be not in the null cone. Then the algorithm outputs $g \in G$ such that $F(g) \leqslant \inf_{g \in G} F(g) + \varepsilon$ within $T = \mathrm{poly}(\log \frac{1}{\varepsilon}, \text{input size}, \frac{1}{\gamma})$ steps.

Analysis: Complexity depends on neighborhood size and diameter bound. Former is controlled by weight norm $L$, latter by inverse weight margin $\frac{1}{\gamma}$.

**State of the art:** Two general algorithms for geodesic convex optimization, which can solve norm minimization, scaling, null cone. Polynomial time for many interesting actions – but *not* always!

# Second order algorithm for norm minimization

Rough idea: Minimize local quadratic approximation (after regularization)

$$Q(H) = F(g) + \nabla F(g)[H] + \frac{1}{2}\nabla^2 F(g)[H,H] \;\;\approx\;\; F(e^H g)$$

on small neighborhoods, where it can be trusted. Need $F$ “robust”.

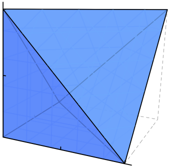
**Theorem**

Let $v \in V$ be not in the null cone. Then the algorithm outputs $g \in G$ such that $F(g) \leqslant \inf_{g \in G} F(g) + \varepsilon$ within $T = \mathrm{poly}(\log \frac{1}{\varepsilon}, \text{input size}, \frac{1}{\gamma})$ steps.

Analysis: Complexity depends on neighborhood size and diameter bound. Former is controlled by weight norm $L$, latter by inverse weight margin $\frac{1}{\gamma}$.

**State of the art:** Two general algorithms for geodesic convex optimization, which can solve norm minimization, scaling, null cone. Polynomial time for many interesting actions – but *not* always!

# Polytopes



# Moment maps and polytopes

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

- depending on the action, $\mu = 0$ means *doubly stochastic* matrix, *trace-preserving and unital* map, ..., uniform marginals

Nonuniform scaling problem:

Given $v \in V$ and $\mathbf{p}$, find $g \in G$ s.th. $\mu(g \cdot v) \approx \mathbf{p}$.

Possible marginals are captured by

$$\Delta(v) = \{\mathbf{p} \ : \ \exists w \in \overline{Gv} : \mu(w) = \mathbf{p}\}$$

- if $G = \mathrm{T}_n$ commutative, simply a Newton polytope [Kostant, Atiyah, ...]
- in general, still convex polytope if defined properly (magically!), but arise *without explicit vertices or facets*! [Kirwan, Mumford, Brion, ...]

# Moment maps and polytopes

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

- depending on the action, $\mu = 0$ means *doubly stochastic* matrix, *trace-preserving and unital* map, . . . , uniform marginals

Nonuniform scaling problem:

$$\text{Given } v \in V \text{ and } \mathbf{p}, \text{ find } g \in G \text{ s.th. } \mu(g \cdot v) \approx \mathbf{p}.$$

Possible marginals are captured by

$$\Delta(v) = \{\mathbf{p} \;:\; \exists w \in \overline{Gv} : \mu(w) = \mathbf{p}\}$$

- if $G = \mathrm{T}_n$ commutative, simply a Newton polytope [Kostant, Atiyah, . . . ]
- in general, still convex polytope if defined properly (magically!), but arise *without explicit vertices or facets*! [Kirwan, Mumford, Brion, . . . ]

# Moment maps and polytopes

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

- depending on the action, $\mu = 0$ means *doubly stochastic* matrix, *trace-preserving and unital* map, ..., uniform marginals

Nonuniform scaling problem:

Given $v \in V$ and $\mathbf{p}$, find $g \in G$ s.th. $\mu(g \cdot v) \approx \mathbf{p}$.

Possible marginals are captured by

$$\Delta(v) = \{\mathbf{p} \ : \ \exists w \in \overline{Gv} : \mu(w) = \mathbf{p}\}$$

- if $G = \mathrm{T}_n$ commutative, simply a Newton polytope [Kostant, Atiyah, ...]
- in general, still convex polytope if defined properly (magically!), but arise *without explicit vertices or facets*! [Kirwan, Mumford, Brion, ...]

# Moment maps and polytopes

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

- depending on the action, $\mu = 0$ means *doubly stochastic* matrix, *trace-preserving and unital* map, …, uniform marginals

Nonuniform scaling problem:

Given $v \in V$ and $\mathbf{p}$, find $g \in G$ s.th. $\mu(g \cdot v) \approx \mathbf{p}$.

Possible marginals are captured by

$$\Delta(v) = \{\mathbf{p} \ : \ \exists w \in \overline{Gv} : \mu(w) = \mathbf{p}\}$$

- if $G = \mathrm{T}_n$ commutative, simply a Newton polytope [Kostant, Atiyah, …]
- in general, still convex polytope if defined properly (magically!), but arise *without explicit vertices or facets*! [Kirwan, Mumford, Brion, …]

# Moment maps and polytopes

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

- depending on the action, $\mu = 0$ means *doubly stochastic* matrix, *trace-preserving and unital* map, . . . , uniform marginals

Nonuniform scaling problem:

Given $v \in V$ and $\mathbf{p}$, find $g \in G$ s.th. $\mu(g \cdot v) \approx \mathbf{p}$.

Possible marginals are captured by moment polytope:

$$\Delta(v) = \{\mathbf{p} \;:\; \exists w \in \overline{Gv} : \mu(w) = \mathbf{p}\}$$

- if $G = \mathrm{T}_n$ commutative, simply a Newton polytope [Kostant, Atiyah, . . . ]
- in general, still convex polytope if defined properly (magically!), but arise *without explicit vertices or facets*! [Kirwan, Mumford, Brion, . . . ]

# Examples of moment polytopes

$G$ commutative (easy):

- Matrix scaling: $\Delta = \{(r, c) : \exists \text{ scaling of } M\} \subseteq \mathbb{R}^{2n}$.
- Schur-Horn: $\Delta = \{\text{diagonal of Hermitian matrix with eigenvalues } \boldsymbol{\lambda}\}$.

$G$ noncommutative (difficult):

- Horn:

  $$\Delta = \{(\lambda_A, \lambda_B, \lambda_C) : A + B = C\} \subseteq \mathbb{R}^{3n}$$

  Complete set of linear inequalities known [Horn, Klyachko, Knutson-Tao, ...].
  Membership in polynomial time, nonuniform scaling open [Mulmuley, Franks].
- Brascamp-Lieb: Validity of integral inequalities in analysis.
- Quantum marginals: What marginals arise by scaling q. states?
  Applications in quantum information, algebraic complexity, algebra...

Typically exponentially many vertices and facets, but succinctly encoded by group action! Which polytopes are captured in this way?

# Examples of moment polytopes

$G$ commutative (easy):

- Matrix scaling: $\Delta = \{(r, c) : \exists \text{ scaling of } M\} \subseteq \mathbb{R}^{2n}$.
- Schur-Horn: $\Delta = \{\text{diagonal of Hermitian matrix with eigenvalues } \boldsymbol{\lambda}\}$.

$G$ noncommutative (difficult):

- Horn:

  $$\Delta = \{(\boldsymbol{\lambda}_A, \boldsymbol{\lambda}_B, \boldsymbol{\lambda}_C) : A + B = C\} \subseteq \mathbb{R}^{3n}$$

  Complete set of linear inequalities known [Horn, Klyachko, Knutson-Tao, ...].
  Membership in polynomial time, nonuniform scaling open [Mulmuley, Franks].
- Brascamp-Lieb: Validity of integral inequalities in analysis.
- Quantum marginals: What marginals arise by scaling q. states?
  Applications in quantum information, algebraic complexity, algebra...

Typically exponentially many vertices and facets, but succinctly encoded by group action! Which polytopes are captured in this way?

# Examples of moment polytopes

$G$ commutative (easy):

- Matrix scaling: $\Delta = \{(r, c) : \exists \text{ scaling of } M\} \subseteq \mathbb{R}^{2n}$.
- Schur-Horn: $\Delta = \{\text{diagonal of Hermitian matrix with eigenvalues } \boldsymbol{\lambda}\}$.

$G$ noncommutative (difficult):

- Horn:

  $$\Delta = \{(\boldsymbol{\lambda}_A, \boldsymbol{\lambda}_B, \boldsymbol{\lambda}_C) : A + B = C\} \subseteq \mathbb{R}^{3n}$$

  Complete set of linear inequalities known [Horn, Klyachko, Knutson-Tao, . . . ].
  Membership in polynomial time, nonuniform scaling open [Mulmuley, Franks].
- Brascamp-Lieb: Validity of integral inequalities in analysis.
- Quantum marginals: What marginals arise by scaling q. states?
  Applications in quantum information, algebraic complexity, algebra. . .

Typically exponentially many vertices and facets, but succinctly encoded by group action! Which polytopes are captured in this way?

# Examples of moment polytopes

$G$ commutative (easy):

- Matrix scaling: $\Delta = \{(r, c) : \exists \text{ scaling of } M\} \subseteq \mathbb{R}^{2n}$.
- Schur-Horn: $\Delta = \{\text{diagonal of Hermitian matrix with eigenvalues } \boldsymbol{\lambda}\}$.

$G$ noncommutative (difficult):

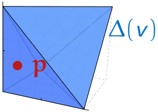
- Horn:

$$\Delta = \{(\boldsymbol{\lambda}_A, \boldsymbol{\lambda}_B, \boldsymbol{\lambda}_C) : A + B = C\} \subseteq \mathbb{R}^{3n}$$

  Complete set of linear inequalities known [Horn, Klyachko, Knutson-Tao, …].
  Membership in polynomial time, nonuniform scaling open [Mulmuley, Franks].
- Brascamp-Lieb: Validity of integral inequalities in analysis.
- Quantum marginals: What marginals arise by scaling q. states?
  Applications in quantum information, algebraic complexity, algebra…

Typically exponentially many vertices and facets, but succinctly encoded by group action! Which polytopes are captured in this way?

# Moment polytopes and noncommutat. group optimization



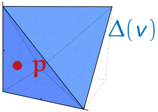
Given $v$ and $\mathbf{p}$, find $g \in G$ such that $\mu(g \cdot v) \approx \mathbf{p}$.

*Key idea:* Reduce to $\mathbf{p} = 0$ by a "shifting trick":

- Laurent polynomials: simply shift exponents $\omega \mapsto \omega - \mathbf{p}$
- If $G$ noncommutative, more involved $V \mapsto V_{\mathbf{p}}$

**State of the art:** Either algorithm discussed above can solve nonuniform scaling problem. Polynomial dependence on most parameters for many interesting actions – but *exponential* dependence on bitsize of $\mathbf{p}$!

# Moment polytopes and noncommutat. group optimization



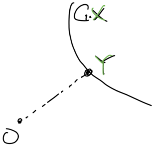
Given $v$ and $\mathbf{p}$, find $g \in G$ such that $\mu(g \cdot v) \approx \mathbf{p}$.

*Key idea:* Reduce to $\mathbf{p} = 0$ by a "shifting trick":

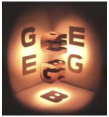
- Laurent polynomials: simply shift exponents $\omega \mapsto \omega - \mathbf{p}$
- If $G$ noncommutative, more involved $V \mapsto V_{\mathbf{p}}$

**State of the art:** Either algorithm discussed above can solve nonuniform scaling problem. Polynomial dependence on most parameters for many interesting actions – but *exponential* dependence on bitsize of $\mathbf{p}$!

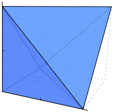
# Summary and outlook



Effective optimization algorithms for $g \mapsto \|g \cdot v\|$, enabled by geodesic convexity and geometric invariant theory.



Moment maps (gradient) capture natural scaling and marginal problems involving probability distributions, quantum states, isotropic position. . . with many applications.



Moment polytopes encode answers to these problems. Often exp. many facets, yet can admit efficient algorithms.

*Many exciting open questions:* Poly-time algorithms for general actions? Better tools for geodesic convex optimization in nonpositive curvature? Structured tensors? Tackle other problems with natural symmetries?

**Thank you for your attention!**