Hidden Symmetries II: Noncommutative Duality, Geodesic Convexity, Polytopes

Michael Walter (University of Amsterdam)

based on joint works with Peter Bürgisser, Levent Dogan, Cole Franks, Ankit Garg, Visu Makam, Harold Nieuwboer, Rafael Oliveira, Avi Wigderson

Geometric Methods in Optimization and Sampling Boot Camp
Simons Institute, Berkeley 2021

QuSoft Research Center for Quantum Software

NWO

Recap of Part I

Alternating minimization – one algorithm that solves two problems:

- *matrix, operator, tensor scaling* with many applications
- *null cone membership* in invariant theory: $0 \in \overline{Gv}$?

Hidden symmetries: Algorithm moves inside group $G = G_1 \times \cdots \times G_d$.
Invariants key to analysis (permanent, Ω-process, ...).

Three questions:

- Why should a simple "greedy" algorithm work?
- What is the connection between scaling and null cone?
- How to go beyond multilinear actions of product groups?

E.g., simultaneous *conjugation*, *symmetric* tensor scaling, ...

Recap of Part I

Alternating minimization – one algorithm that solves two problems:

- *matrix, operator, tensor scaling* with many applications
- *null cone membership* in invariant theory: $0 \in \overline{Gv}$?

Hidden symmetries: Algorithm moves inside group $G = G_1 \times \cdots \times G_d$.
Invariants key to analysis (permanent, Ω-process, ...).

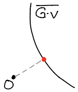
Three questions:

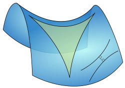
- Why should a simple "greedy" algorithm work?
- What is the connection between scaling and null cone?
- How to go beyond multilinear actions of product groups?

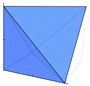
E.g., simultaneous *conjugation*, *symmetric* tensor scaling, ...

Plan for Part II

1. Group actions and optimization
2. From Euclidean to geodesic convexity
3. Noncommutative duality
4. Algorithms for geodesic optimization
5. Polytopes and nonuniform scaling

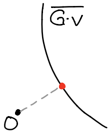






Big picture: Null cone, optimization, and scaling

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

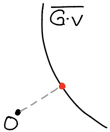


Big picture: Null cone, optimization, and scaling

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

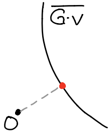
$\Updownarrow$

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

$\Uparrow$

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

Norm Minimization

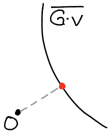


Big picture: Null cone, optimization, and scaling

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

$\Updownarrow$

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



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

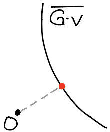
Norm Minimization — Scaling Problem

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. $\nabla_g \|g \cdot v\| \approx 0$.

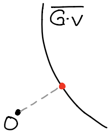
Norm Minimization | Scaling Problem

Big picture: Null cone, optimization, and scaling

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

$\Updownarrow$

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

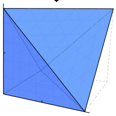


Minimize $\|g \cdot v\|$ over $g \in G$. $\xLeftrightarrow[\text{convexity}]{\text{geodesic}}$ Find $g \in G$ s.th. $\nabla_g \|g \cdot v\| \approx 0$.

Norm Minimization

Scaling Problem

$\Updownarrow$



Polytopes

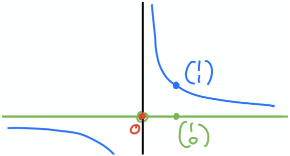
Let's get started!

Setup

Group $G \subseteq \mathrm{GL}_n(\mathbb{C})$ "nice" (reductive), such as GL_n, SL_n, or $\mathrm{T}_n = (\mathbb{C}^*)^n$
Action on $V = \mathbb{C}^m$ by linear transformations
Orbits $Gv = \{g \cdot v \,:\, g \in G\}$ and their closures $\overline{Gv}$

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

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



The minimum ℓ^2-norm in an orbit closure is called the capacity:

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

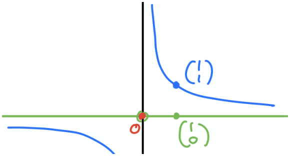
- the basic optimization problem that we wish to solve!
- generalizes notions of matrix, polynomial, operator capacity

Setup

Group $G \subseteq \mathrm{GL}_n(\mathbb{C})$ "nice" (reductive), such as GL_n, SL_n, or $\mathrm{T}_n = (\mathbb{C}^*)^n$
Action on $V = \mathbb{C}^m$ by linear transformations
Orbits $Gv = \{g \cdot v \ : \ g \in G\}$ and their closures $\overline{Gv}$

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

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



The minimum ℓ^2-norm in an orbit closure is called the capacity:

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

- the basic optimization problem that we wish to solve!
- generalizes notions of matrix, polynomial, operator capacity

Setup

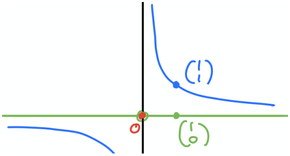
Group $G \subseteq \mathrm{GL}_n(\mathbb{C})$ "nice" (reductive), such as GL_n, SL_n, or $\mathrm{T}_n = (\mathbb{C}^*)^n$

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

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

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

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

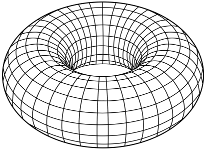


The minimum ℓ^2-norm in an orbit closure is called the capacity:

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

- the basic optimization problem that we wish to solve!
- generalizes notions of matrix, polynomial, operator capacity

Warmup: Commutative Group Actions



Example: Matrix scaling revisited

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

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

Capacity:

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

- geometric program, log-convex in x, y

Gradient:

$$\nabla_{x=y=0} \log(\dots) = \big(\mathbf{r}(M), \mathbf{c}(M)\big)$$

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

Norm minimization and matrix scaling are equivalent! ☺ Motivates why Sinkhorn solves either and is starting point for cutting-edge algorithms.

Example: Matrix scaling revisited

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

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

Capacity:

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

- geometric program, log-convex in x, y

Gradient:

$$\nabla_{x=y=0} \log(\dots) = (\mathbf{r}(M), \mathbf{c}(M))$$

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

Norm minimization and matrix scaling are equivalent! ☺ Motivates why Sinkhorn solves either and is starting point for cutting-edge algorithms.

Example: Matrix scaling revisited

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

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

Capacity:

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

- geometric program, log-convex in x, y

Gradient:

$$\nabla_{x=y=0} \log(\dots) = (\mathbf{r}(M), \mathbf{c}(M))$$

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

Norm minimization and matrix scaling are equivalent! ☺ Motivates why Sinkhorn solves either and is starting point for cutting-edge algorithms.

Example: Matrix scaling revisited

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

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

Capacity:

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

- geometric program, log-convex in x, y

Gradient:

$$\nabla_{x=y=0} \log(\dots) = \big(\mathbf{r}(M), \mathbf{c}(M)\big)$$

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

Norm minimization and matrix scaling are equivalent! ☺ Motivates why Sinkhorn solves either and is starting point for cutting-edge algorithms.

Example: Matrix scaling revisited

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

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

Capacity:

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

- geometric program, log-convex in x, y

Gradient:

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

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

Norm minimization and matrix scaling are equivalent! ☺ Motivates why Sinkhorn solves either and is starting point for cutting-edge algorithms.

Example: Laurent polynomials

$G = \mathrm{T}_n(\mathbb{C})$ acts on Laurent polynomials in n variables by scaling:

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

Capacity:

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

- again geometric program, log-convex in x
- $\mathrm{cap}(P) > 0 \quad \text{iff} \quad 0 \in$ Newton polytope of P

Gradient:

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

Any action of T_n is essentially of this form. Rich and nontrivial!

Example: Laurent polynomials

$G = T_n(\mathbb{C})$ acts on Laurent polynomials in n variables by scaling:

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

Capacity:

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

- again geometric program, log-convex in x
- $\mathrm{cap}(P) > 0$ iff $0 \in$ Newton polytope of P

Gradient:

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

Any action of T_n is essentially of this form. Rich and nontrivial!

Example: Laurent polynomials

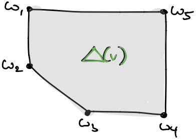
$G = T_n(\mathbb{C})$ acts on Laurent polynomials in n variables by scaling:

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

Capacity:

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

- again geometric program, log-convex in x
- $\operatorname{cap}(P) > 0$ iff $0 \in$ Newton polytope of P



Gradient:

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

Any action of T_n is essentially of this form. Rich and nontrivial!

Example: Laurent polynomials

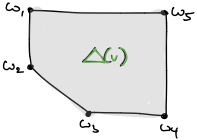
$G = \mathrm{T}_n(\mathbb{C})$ acts on Laurent polynomials in n variables by scaling:

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

Capacity:

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

- again geometric program, log-convex in x
- $\operatorname{cap}(P) > 0$ iff $0 \in$ Newton polytope of P

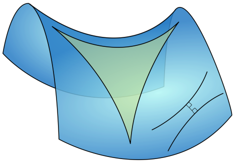


Gradient:

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

Any action of T_n is essentially of this form. Rich and nontrivial!

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:

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

The gradient after this change of variables is called the moment map:

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

- Riemannian gradient at $p = I$, as a function of v
- 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:

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

The gradient after this change of variables is called the moment map:

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

- Riemannian gradient at $p = I$, as a function of v
- 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:

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

The gradient after this change of variables is called the moment map:

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

- Riemannian gradient at $p = I$, as a function of v
- 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:

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

The gradient after this change of variables is called the moment map:

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

- Riemannian gradient at $p = I$, as a function of v
- 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 $G = \mathrm{GL}_n(\mathbb{C}) \times \mathrm{GL}_n(\mathbb{C})$ act on $V = \mathrm{Mat}_n(\mathbb{C})^{\oplus d}$ by left-right action:

$$(g, h) \cdot (M_1, \ldots, M_d) = (gM_1 h^{-1}, \ldots, gM_d h^{-1})$$

Capacity:

$$\mathrm{cap}(M)^2 = \inf_{g,h} \sum_{i=1}^{d} \|gM_i h^{-1}\|_F^2 = \inf_{X,Y \in \mathrm{Herm}_n} \sum_{i=1}^{d} \mathrm{tr}\, e^X M_i e^{-Y} M_i^*$$

Moment map (gradient):

$$\mu(M) = \frac{1}{\|M\|^2} \left(\sum_{i=1}^{d} M_i M_i^*, -\sum_{i=1}^{d} M_i^* M_i \right)$$

If we restrict 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 $G = \mathrm{GL}_n(\mathbb{C}) \times \mathrm{GL}_n(\mathbb{C})$ act on $V = \mathrm{Mat}_n(\mathbb{C})^{\oplus d}$ by left-right action:

$$(g, h) \cdot (M_1, \ldots, M_d) = (gM_1 h^{-1}, \ldots, gM_d h^{-1})$$

Capacity:

$$\mathrm{cap}(M)^2 = \inf_{g,h} \sum_{i=1}^{d} \|gM_i h^{-1}\|_F^2 = \inf_{X,Y \in \mathrm{Herm}_n} \sum_{i=1}^{d} \operatorname{tr} e^X M_i e^{-Y} M_i^*$$

Moment map (gradient):

$$\mu(M) = \frac{1}{\|M\|^2} \left(\sum_{i=1}^{d} M_i M_i^*, -\sum_{i=1}^{d} M_i^* M_i \right)$$

If we restrict 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 $G = \mathrm{GL}_n(\mathbb{C}) \times \mathrm{GL}_n(\mathbb{C})$ act on $V = \mathrm{Mat}_n(\mathbb{C})^{\oplus d}$ by left-right action:

$$(g, h) \cdot (M_1, \ldots, M_d) = (gM_1 h^{-1}, \ldots, gM_d h^{-1})$$

Capacity:

$$\mathrm{cap}(M)^2 = \inf_{g,h} \sum_{i=1}^{d} \|gM_i h^{-1}\|_F^2 = \inf_{X,Y \in \mathrm{Herm}_n} \sum_{i=1}^{d} \mathrm{tr}\, e^{X} M_i e^{-Y} M_i^*$$

Moment map (gradient):

$$\mu(M) = \frac{1}{\|M\|^2}\left(\sum_{i=1}^{d} M_i M_i^*, -\sum_{i=1}^{d} M_i^* M_i\right)$$

If we restrict 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 $G = \mathrm{GL}_n(\mathbb{C}) \times \mathrm{GL}_n(\mathbb{C})$ act on $V = \mathrm{Mat}_n(\mathbb{C})^{\oplus d}$ by left-right action:

$$(g, h) \cdot (M_1, \ldots, M_d) = (gM_1 h^{-1}, \ldots, gM_d h^{-1})$$

Capacity:

$$\mathrm{cap}(M)^2 = \inf_{g,h} \sum_{i=1}^{d} \|gM_i h^{-1}\|_F^2 = \inf_{X,Y \in \mathrm{Herm}_n} \sum_{i=1}^{d} \mathrm{tr}\, e^X M_i e^{-Y} M_i^*$$

Moment map (gradient):

$$\mu(M) = \frac{1}{\|M\|^2} \left(\sum_{i=1}^d M_i M_i^*, -\sum_{i=1}^d M_i^* M_i \right)$$

If we restrict 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 $G = \mathrm{GL}_n(\mathbb{C}) \times \mathrm{GL}_n(\mathbb{C})$ act on $V = \mathrm{Mat}_n(\mathbb{C})^{\oplus d}$ by left-right action:

$$(g, h) \cdot (M_1, \ldots, M_d) = (gM_1 h^{-1}, \ldots, gM_d h^{-1})$$

Capacity:

$$\mathrm{cap}(M)^2 = \inf_{g,h} \sum_{i=1}^{d} \|gM_i h^{-1}\|_F^2 = \inf_{X,Y \in \mathrm{Herm}_n} \sum_{i=1}^{d} \mathrm{tr}\, e^X M_i e^{-Y} M_i^*$$

Moment map (gradient):

$$\mu(M) = \frac{1}{\|M\|^2} \left(\sum_{i=1}^{d} M_i M_i^*, -\sum_{i=1}^{d} M_i^* M_i \right)$$

If we restrict to $G = \mathrm{SL}_n \times \mathrm{SL}_n$: captures operator scaling!

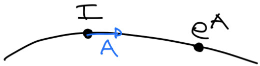
However, the objective is *not* convex in X, Y...

Geodesic convexity

Why does the equivalence between norm minimization and scaling hold?

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

is convex along the curves e^{Xt} for any $X \in \mathrm{Herm}_n$, which are geodesics for a natural Riemannian metric on PD_n. That is, F is geodesically convex!



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

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

$$\|v\| = \mathrm{cap}(v) \quad \Leftrightarrow \quad \mu(v) = 0$$

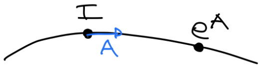
How convex for given action? Necessary for algorithms!

Geodesic convexity

Why does the equivalence between norm minimization and scaling hold?

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

is convex along the curves e^{Xt} for any $X \in \mathrm{Herm}_n$, which are geodesics for a natural Riemannian metric on PD_n. That is, F is geodesically convex!



Proof? $\{e^{Xt}\}$ = commutative subgroup ⇒ Laurent polynomials ☺

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

$$\|v\| = \mathrm{cap}(v) \qquad \Leftrightarrow \qquad \mu(v) = 0$$

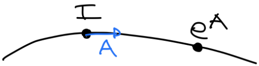
How convex for given action? Necessary for algorithms!

Geodesic convexity

Why does the equivalence between norm minimization and scaling hold?

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

is convex along the curves e^{Xt} for any $X \in \mathrm{Herm}_n$, which are geodesics for a natural Riemannian metric on PD_n. That is, F is geodesically convex!



Proof? $\{e^{Xt}\}$ = commutative subgroup ⇒ Laurent polynomials ☺

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

$$\|v\| = \mathrm{cap}(v) \qquad \Leftrightarrow \qquad \mu(v) = 0$$

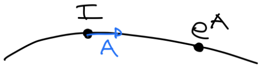
How convex for given action? Necessary for algorithms!

Geodesic convexity

Why does the equivalence between norm minimization and scaling hold?

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

is convex along the curves e^{Xt} for any $X \in \mathrm{Herm}_n$, which are geodesics for a natural Riemannian metric on PD_n. That is, F is geodesically convex!



Proof? $\{e^{Xt}\}$ = commutative subgroup ⇒ Laurent polynomials ☺

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

$$\|v\| = \mathrm{cap}(v) \qquad \Leftrightarrow \qquad \mu(v) = 0$$

How convex for given action? Necessary for algorithms!

Geodesic convexity made quantitative

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

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

Noncommutative duality estimates

$$1 - \frac{\|\mu(v)\|}{\gamma} \leqslant \frac{\operatorname{cap}(v)^2}{\|v\|^2} \leqslant 1 - \frac{\|\mu(v)\|^2}{2L}$$

- ☺ 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, γ depend on combinatorial data of action.

Geodesic convexity made quantitative

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

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

Noncommutative duality estimates

$$1 - \frac{\|\mu(v)\|}{\gamma} \leqslant \frac{\operatorname{cap}(v)^2}{\|v\|^2} \leqslant 1 - \frac{\|\mu(v)\|^2}{2L}$$

- ☺ 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, γ depend on combinatorial data of action.

Noncommutative duality

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

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

Noncommutative duality estimates

$$1 - \frac{\|\mu(v)\|}{\gamma} \leqslant \frac{\mathrm{cap}(v)^2}{\|v\|^2} \leqslant 1 - \frac{\|\mu(v)\|^2}{2L}$$

- ☺ 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, γ depend on combinatorial data of action.

Noncommutative duality

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

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

Noncommutative duality estimates

$$1 - \frac{\|\mu(v)\|}{\gamma} \leqslant \frac{\mathrm{cap}(v)^2}{\|v\|^2} \leqslant 1 - \frac{\|\mu(v)\|^2}{2L}$$

- ☺ 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]

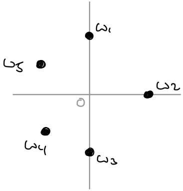
Parameters L, γ depend on combinatorial data of action.

Interlude: Weights of action

Take any action of GL_n. If we restrict to $\mathrm{T}_n = \begin{pmatrix} \cdot & \\ & \cdot_{\cdot} \end{pmatrix}$, 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.



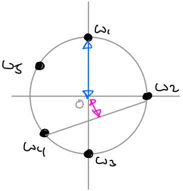
Their geometry determine the geodesic convexity parameters L, γ.

Interlude: Weights of action

Take any action of GL_n. If we restrict to $\mathrm{T}_n = \begin{pmatrix} \cdot & \\ & \cdot_{\cdot} \end{pmatrix}$, 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.

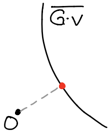


Their geometry determine the geodesic convexity parameters L, γ.

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 $g \in G$. $\quad \xLeftarrow[\text{NC-duality}]{\text{g-convexity}} \!\!\Rightarrow \quad$ 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 (not) discussed in Avi' talk fall into this framework.

Interlude: Beyond GL_n and SL_n

All the preceding generalizes to complex reductive groups – not just SL_n, T_n, ST_n, and products thereof. Concretely, this means a subgroup

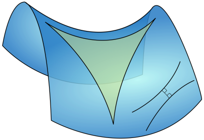
$$G \subseteq \mathsf{GL}_n(\mathbb{C})$$

defined by polynomial equations that is closed under taking adjoints.

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

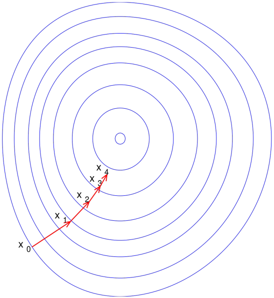
$$G \cap \mathsf{PD}_n = \{g^* g : g \in G\}.$$

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



NB: 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 (Nicholas's talk). Combine with a priori lower bound obtained using constructive invariant theory (Avi's talk).

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 (Nicholas's talk). Combine with a priori lower bound obtained using constructive invariant theory (Avi's talk).

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 (Nicholas's talk). Combine with a priori lower bound obtained using constructive invariant theory (Avi's talk).

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)

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

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 $T = \text{poly}(\log \frac{1}{\varepsilon}, \text{input size}, \frac{1}{\gamma})$ 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".

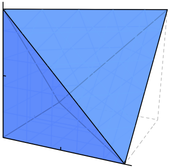
Theorem

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, *isotropic* frame, ..., 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 $G = T_n$ 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, *isotropic* frame, ..., 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 $G = T_n$ 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, *isotropic* frame, ..., 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, *isotropic* frame, ..., 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, *isotropic* frame, ..., 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 tensors?
  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 tensors?
  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 tensors? 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):

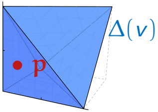
- Horn:

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

  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 tensors?
  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 noncommutative optimization



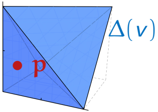
Given v and $\mathbf{p}$, find $g \in G$ such that $\mu(g \cdot v) \approx \mathbf{p}$.

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 noncommutative optimization



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

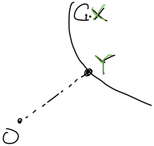
Key idea: Reduce to $\mathbf{p} = 0$ by a "shifting trick":

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

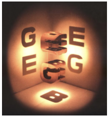
State of the art: Either algorithm discussed above can solve nonuniform scaling problem. Polynomial dependence on most parameters for many interesting actions – but *exponential* dependence on bitsize of $\mathbf{p}$!

Summary and outlook

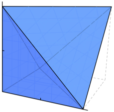
Geodesic convexity of $g \mapsto \|g \cdot v\|$ underlies unreasonable effectiveness of alternating minimization, is key to general efficient algorithms that exploit hidden symmetries.



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? What is tractable in invariant theory? How to tackle other difficult problems with natural symmetries? **Thank you for your attention!**