Symmetries of Computational Problems & Optimization

Michael Walter (Uni Bochum)

Abel Prize Laureates Lectures, Amsterdam, April 2022

joint works with P. Bürgisser, L. Dogan, C. Franks, A. Garg, V. Makam, H. Nieuwboer, R. Oliveira, A. Ramachandran, **Avi Wigderson**

RUHR UNIVERSITÄT BOCHUM RUB
CASA CYBER SECURITY IN THE AGE OF LARGE-SCALE ADVERSARIES
erc European Research Council Established by the European Commission
UNIVERSITY OF AMSTERDAM
QuSoft Research Center for Quantum Software

Prelude: Matrix scaling

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

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

A matrix is called *doubly stochastic* if row & column sums are 1.

Matrix scaling problem: Given X, find approx. doubly stochastic scalings.

Sinkhorn algorithm: 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, and in polynomial time! [LSW]
- Possible iff $\operatorname{per}(X) > 0$ iff bipartite perfect matching in support of X.

Applications to statistics, machine learning, complexity, combinatorics, numerics, ...

Prelude: Matrix scaling

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

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

A matrix is called *doubly stochastic* if row & column sums are 1.

Matrix scaling problem: Given X, find approx. doubly stochastic scalings.

Sinkhorn algorithm: 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, and in polynomial time! [LSW]
- Possible iff $\mathrm{per}(X) > 0$ iff bipartite perfect matching in support of X.

Applications to statistics, machine learning, complexity, combinatorics, numerics, ...

Prelude: Matrix scaling

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

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

A matrix is called *doubly stochastic* if row & column sums are 1.

Matrix scaling problem: Given X, find approx. doubly stochastic scalings.

Sinkhorn algorithm: 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, and in polynomial time! [LSW]
- Possible iff $\mathrm{per}(X) > 0$ iff bipartite perfect matching in support of X.

Applications to statistics, machine learning, complexity, combinatorics, numerics, . . .

Prelude: Matrix scaling

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

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

A matrix is called *doubly stochastic* if row & column sums are 1.

Matrix scaling problem: Given X, find approx. doubly stochastic scalings.

Sinkhorn algorithm: Alternatingly normalize rows & columns:

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

- This converges whenever possible, and in polynomial time! [LSW]
- Possible iff $\mathrm{per}(X) > 0$ iff bipartite perfect matching in support of X.

Applications to statistics, machine learning, complexity, combinatorics, numerics, ...

Prelude: Matrix scaling

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

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

A matrix is called *doubly stochastic* if row & column sums are 1.

Matrix scaling problem: Given X, find approx. doubly stochastic scalings.

Sinkhorn algorithm: 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, and in polynomial time! [LSW]
- Possible iff $\text{per}(X) > 0$ iff bipartite perfect matching in support of X.

Applications to statistics, machine learning, complexity, combinatorics, numerics, . . .

Prelude: Matrix scaling

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

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

A matrix is called *doubly stochastic* if row & column sums are 1.

Matrix scaling problem: Given X, find approx. doubly stochastic scalings.

Sinkhorn algorithm: Alternatingly normalize rows & columns:

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

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

Overview

A series of recent works discovered clues that hidden symmetries and optimization connect a wide range of problems:

program testing | tensor isomorphism | quantum entanglement

P vs NP | algebraic curves | ··· | quantum materials

derandomization | optimal transport | maximum likelihood

This discovery was already key to fast algorithms and structural insight.

Plan for today: Introduction to these connections, some applications, and a glance of how optimization in curved spaces can lead to progress.

Overview

A series of recent works discovered clues that hidden symmetries and optimization connect a wide range of problems:

theoretical program testing

tensor isomorphism

quantum entanglement

P vs NP

algebraic curves

· · ·

quantum materials

derandomization

optimal transport

maximum likelihood

This discovery was already key to fast algorithms and structural insight.

Plan for today: Introduction to these connections, some applications, and a glance of how optimization in curved spaces can lead to progress.

Overview

A series of recent works discovered clues that hidden symmetries and optimization connect a wide range of problems:

theoretical
program testing

algebraic
tensor isomorphism

quantum entanglement

P vs NP

algebraic curves

· · ·

quantum materials

derandomization

optimal transport

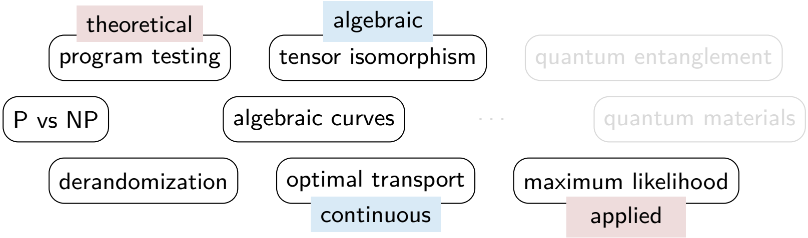
maximum likelihood

This discovery was already key to fast algorithms and structural insight.

Plan for today: Introduction to these connections, some applications, and a glance of how optimization in curved spaces can lead to progress.

Overview

A series of recent works discovered clues that hidden symmetries and optimization connect a wide range of problems:

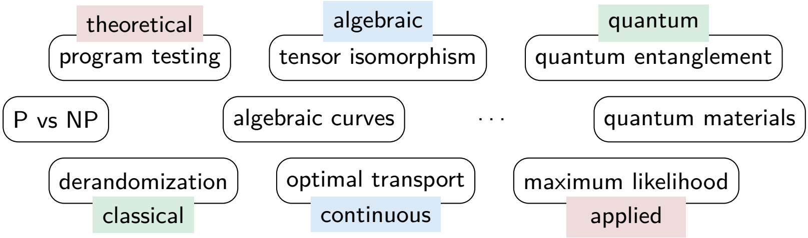


This discovery was already key to fast algorithms and structural insight.

Plan for today: Introduction to these connections, some applications, and a glance of how optimization in curved spaces can lead to progress.

Overview

A series of recent works discovered clues that hidden symmetries and optimization connect a wide range of problems:

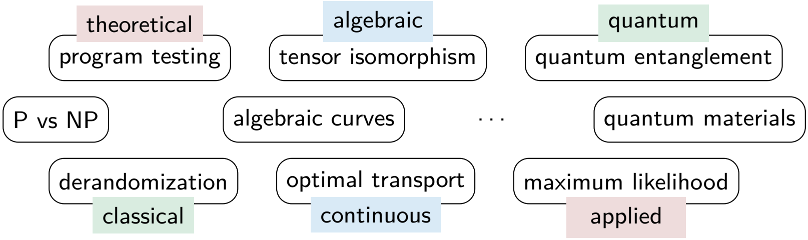


This discovery was already key to fast algorithms and structural insight.

Plan for today: Introduction to these connections, some applications, and a glance of how optimization in curved spaces can lead to progress.

Overview

A series of recent works discovered clues that hidden symmetries and optimization connect a wide range of problems:

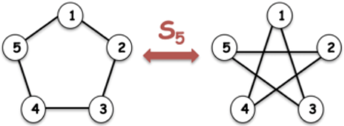


This discovery was already key to fast algorithms and structural insight.

Plan for today: Introduction to these connections, some applications, and a glance of how optimization in curved spaces can lead to progress.

Symmetries and group actions

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



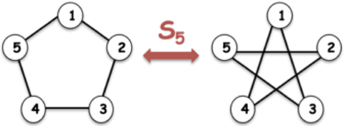
Problem: How can we algorithmically and efficiently check equivalence?

Interesting (and often difficult) problems with many applications:

- no polynomial-time algorithm known for graph isomorphism
- matrices equivalent iff equal rank, but how about tensors?
- derandomizing polynomial identity testing implies circuit lower bounds
- computing *normal forms*, describing *moduli spaces* and *invariants*. . .

Symmetries and group actions

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



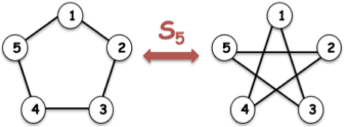
Problem: How can we algorithmically and efficiently check equivalence?

Interesting (and often difficult) problems with many applications:

- no polynomial-time algorithm known for graph isomorphism
- matrices equivalent iff equal rank, but how about tensors?
- derandomizing polynomial identity testing implies circuit lower bounds
- computing *normal forms*, describing *moduli spaces* and *invariants*...

Symmetries and group actions

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



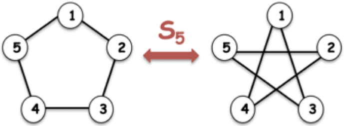
Problem: How can we algorithmically and efficiently check equivalence?

Interesting (and often difficult) problems with many applications:

- no polynomial-time algorithm known for graph isomorphism
- matrices equivalent iff equal rank, but how about tensors?
- derandomizing polynomial identity testing implies circuit lower bounds
- computing *normal forms*, describing *moduli spaces* and *invariants*. . .

Symmetries and group actions

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



Problem: How can we algorithmically and efficiently check equivalence?

Interesting (and often difficult) problems with many applications:

- no polynomial-time algorithm known for graph isomorphism
- matrices equivalent iff equal rank, but how about tensors?
- derandomizing polynomial identity testing implies circuit lower bounds
- computing *normal forms*, describing *moduli spaces* and *invariants*. . .

Orbit problems

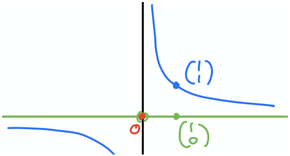
Group $G \subseteq \mathrm{GL}_n(\mathbb{C})$ "nice", such as GL_n, SL_n, or $\mathrm{T}_n = \left(\ddots\right)$

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

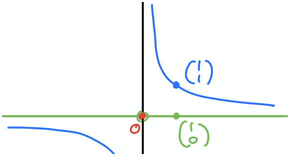
Group $G \subseteq \mathrm{GL}_n(\mathbb{C})$ "nice", such as GL_n, SL_n, or $\mathrm{T}_n = \left(\ddots \right)$

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

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

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

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



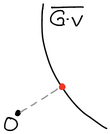
Orbit problems:

- 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:** $0 \in \overline{Gv}$?

Big picture: Null cone, optimization, and scaling

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

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



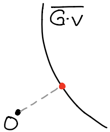
Big picture: Null cone, optimization, and scaling

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

Algebra

$\Updownarrow$

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



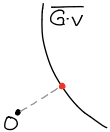
Big picture: Null cone, optimization, and scaling

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

Algebra

$\Updownarrow$

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



$\Nearrow$

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

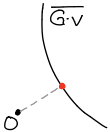
Optimization

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

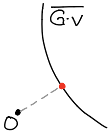
Optimization — Scaling Problem

Big picture: Null cone, optimization, and scaling

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

$\Updownarrow$

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



$\Uparrow$ (from Optimization)

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

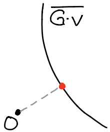
Optimization — Scaling Problem

Big picture: Null cone, optimization, and scaling

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

$\Updownarrow$

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



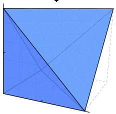
$\Nearrow$

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

Optimization

Scaling Problem

$\Updownarrow$



Polytopes

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

Example: Matrix scaling revisited

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

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

A matrix is called *doubly stochastic* if row & column sums are 1.

Matrix scaling problem: Given X, find approx. doubly stochastic scalings.

Indeed a "scaling problem" in the general sense!

Example: Matrix scaling revisited

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

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

A matrix is called *doubly stochastic* if row & column sums are 1.

Matrix scaling problem: Given X, find approx. doubly stochastic scalings.

Indeed a "scaling problem" in the general sense!

$$V = \mathrm{Mat}_{n\times n}, \quad G = \mathrm{T}_n \times \mathrm{T}_n, \quad (g_1, g_2)v = g_1 v g_2.$$

Then, $\nabla \|g \cdot v\|^2 = (\text{row sums}, \text{column sums})$ of $X_{ij} = |v_{ij}|^2$.

Example: Matrix scaling revisited

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

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

A matrix is called *doubly stochastic* if row & column sums are 1.

Matrix scaling problem: Given X, find approx. doubly stochastic scalings.

Indeed a "scaling problem" in the general sense! ☺ This explains...

- connection between permanent and matrix scaling.
- why Sinkhorn works. starting point for cutting-edge algos.

In general, commutative actions capture linear & geometric programming!

Example: Matrix scaling revisited

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

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

A matrix is called *doubly stochastic* if row & column sums are 1.

Matrix scaling problem: Given X, find approx. doubly stochastic scalings.

Indeed a "scaling problem" in the general sense! ☺ This explains...

- connection between permanent and matrix scaling.
- why Sinkhorn works. starting point for cutting-edge algos.

In general, commutative actions capture linear & 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 ρ, find scaling such that $\sigma_1, \sigma_2 \approx I$.

Tensor scaling problem: Given ρ, which $(\sigma_1, \dots, \sigma_d)$ can be obtained by scaling?

- eigenvalues form *convex polytopes*
- applications in quantum information, algebraic complexity, algebra...
- exp. many vertices and facets, but succinctly encoded by group action

Which other interesting polytopes captured in this way?
Solve more combinatorial problems by optimization?

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 ρ, find scaling such that $\sigma_1, \sigma_2 \approx I$.

Tensor scaling problem: Given ρ, which $(\sigma_1, \dots, \sigma_d)$ can be obtained by scaling?

- eigenvalues form *convex polytopes*
- applications in quantum information, algebraic complexity, algebra...
- exp. many vertices and facets, but succinctly encoded by group action

Which other interesting polytopes captured in this way? Solve more combinatorial problems by optimization?

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 ρ, find scaling such that $\sigma_1, \sigma_2 \approx I$.

Tensor scaling problem: Given ρ, which $(\sigma_1, \dots, \sigma_d)$ can be obtained by scaling?

- eigenvalues form *convex polytopes*
- applications in quantum information, algebraic complexity, algebra...
- exp. many vertices and facets, but succinctly encoded by group action

Which other interesting polytopes captured in this way?
Solve more combinatorial problems by optimization?

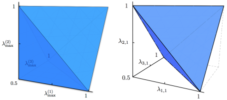
Example: Operator and tensor scaling

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

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

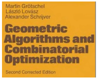
Operator scaling problem: Given ρ, find scaling such that $\sigma_1, \sigma_2 \approx I$.

Tensor scaling problem: Given ρ, which $(\sigma_1, \ldots, \sigma_d)$ can be obtained by scaling?



- eigenvalues form *convex polytopes*
- applications in quantum information, algebraic complexity, algebra...
- exp. many vertices and facets, but succinctly encoded by group action

Which other interesting polytopes captured in this way?
Solve more combinatorial problems by optimization?



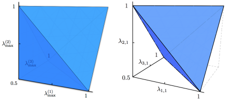
Example: Operator and tensor scaling

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

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

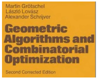
Operator scaling problem: Given ρ, find scaling such that $\sigma_1, \sigma_2 \approx I$.

Tensor scaling problem: Given ρ, which $(\sigma_1, \ldots, \sigma_d)$ can be obtained by scaling?



- eigenvalues form *convex polytopes*
- applications in quantum information, algebraic complexity, algebra...
- exp. many vertices and facets, but succinctly encoded by group action

Which other interesting polytopes captured in this way?
Solve more combinatorial problems by optimization?



Example: Operator scaling and polynomial identity testing

We can identify ρ, σ with 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: $Y_k = g X_k h^T$.

Operator scaling problem: Given Φ, find unital & trace-preserving scaling.

Possible iff $\det \sum_k \alpha_k \otimes X_k \neq 0$ for matrices α_k.

- means symbolic matrix in NC variables α_k has *maximal NC-rank*
- when α_k restricted to scalars: major open problem in TCS!

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

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

Example: Operator scaling and polynomial identity testing

We can identify ρ, σ with 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: $Y_k = g X_k h^T$.

Operator scaling problem: Given Φ, find unital & trace-preserving scaling.

Possible iff $\det \sum_k \alpha_k \otimes X_k \neq 0$ for matrices α_k.

- means symbolic matrix in NC variables α_k has *maximal NC-rank*
- when α_k restricted to scalars: major open problem in TCS!

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

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

Example: Operator scaling and polynomial identity testing

We can identify ρ, σ with 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: $Y_k = g X_k h^T$.

Operator scaling problem: Given Φ, find unital & trace-preserving scaling.

Possible iff $\det \sum_k \alpha_k \otimes X_k \neq 0$ for matrices α_k.

- means symbolic matrix in NC variables α_k has *maximal NC-rank*
- when α_k restricted to scalars: major open problem in TCS!

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

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

Example: Operator scaling and polynomial identity testing

We can identify ρ, σ with 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: $Y_k = g X_k h^T$.

Operator scaling problem: Given Φ, find unital & trace-preserving scaling.

Possible iff $\det \sum_k \alpha_k \otimes X_k \neq 0$ for matrices α_k.

- means symbolic matrix in NC variables α_k has *maximal NC-rank*
- when α_k restricted to scalars: major open problem in TCS!

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

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

Many other connections and applications

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

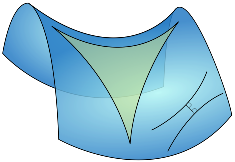
Analysis: Brascamp-Lieb inequalities, solution of Paulsen's problem

Symplectic geometry: Horn's problem $\exists\ A + B = C$ with spectrum α, β, γ?
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
Quantum physics: Tensor network algorithms

Symmetry and Optimization



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, can restrict to:

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

This is a Hadamard manifold, a particularly nice Riemannian manifold of nonpositive curvature.

The gradient $\nabla F(I) = \nabla_{X=0} F(e^X)$ is known as *moment map* in geometry & physics. It turns out $\nabla F = 0$ captures natural scaling problems!

Norm minimization and gradient

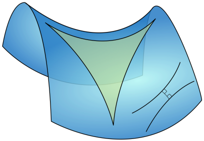
We want to minimize the function:

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

Consider $G = \mathrm{GL}_n$. By the polar decomposition, can restrict to:

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

This is a Hadamard manifold, a particularly nice Riemannian manifold of nonpositive curvature.



The gradient $\nabla F(I) = \nabla_{X=0} F(e^X)$ is known as *moment map* in geometry & physics. It turns out $\nabla F = 0$ captures natural scaling problems!

Norm minimization and gradient

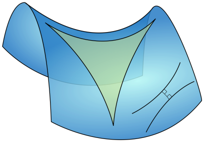
We want to minimize the function:

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

Consider $G = \mathrm{GL}_n$. By the polar decomposition, can restrict to:

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

This is a Hadamard manifold, a particularly nice Riemannian manifold of nonpositive curvature.



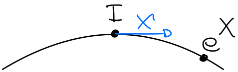
The gradient $\nabla F(I) = \nabla_{X=0} F(e^X)$ is known as *moment map* in geometry & physics. It turns out $\nabla F = 0$ captures natural scaling problems!

Geodesic convexity

While not convex in the usual sense, the objective

$$F(g) = \log \|g \cdot v\|$$

is convex along the geodesics e^{Xt} of PD_n, i.e., $\partial_t^2 F(e^{Xt}) \geqslant 0$. [Kempf-Ness]



Just like in the Euclidean case, this means critical points are global minima.

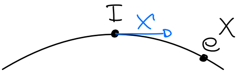
How convex for given action? Necessary for algorithms!

Geodesic convexity

While not convex in the usual sense, the objective

$$F(g) = \log \|g \cdot v\|$$

is convex along the geodesics e^{Xt} of PD_n, i.e., $\partial_t^2 F(e^{Xt}) \geqslant 0$. [Kempf-Ness]



Just like in the Euclidean case, this means critical points are global minima.

How convex for given action? Necessary for algorithms!

Geodesic convexity made quantitative

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

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

Moreover, **noncommutative duality estimates**: For $F_* = \inf_g F(g)$,

$$1 - \frac{\|\nabla F\|}{\gamma} \leqslant e^{F_* - F} \leqslant 1 - \frac{\|\nabla F\|^2}{2L}$$

- ☺ relates norm minimization ⇔ scaling in a quantitative way
- ☺ implies either can solve null cone problem!

Parameters L, γ depend on combinatorial data of action.

Geodesic convexity made quantitative

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

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

Moreover, **noncommutative duality estimates**: For $F_* = \inf_g F(g)$,

$$1 - \frac{\|\nabla F\|}{\gamma} \leqslant e^{F_* - F} \leqslant 1 - \frac{\|\nabla F\|^2}{2L}$$

- ☺ relates norm minimization ⇔ scaling in a quantitative way
- ☺ implies either can solve null cone problem!

Parameters L, γ depend on combinatorial data of action.

Geodesic convexity made quantitative

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

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

Moreover, **noncommutative duality estimates**: For $F_* = \inf_g F(g)$,

$$1 - \frac{\|\nabla F\|}{\gamma} \leqslant e^{F_* - F} \leqslant 1 - \frac{\|\nabla F\|^2}{2L}$$

☺ relates norm minimization ⇔ scaling in a quantitative way

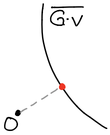
☺ implies either can solve null cone problem!

Parameters L, γ depend on combinatorial data of action.

Framework: Noncommutative group optimization [BFGOWW]

Action of "nice" $G \subseteq \mathrm{GL}_n$ on $V \cong \mathbb{C}^m$.

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



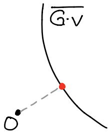
Minimize $\|g \cdot v\|$ over $g \in G$. $\underset{\text{NC-duality}}{\overset{\text{g-convexity}}{\Longleftrightarrow}}$ Find $g \in G$ s.th. $\nabla \|g \cdot v\| \approx 0$.

Norm Minimization — Scaling Problem

- All examples mentioned earlier fall into this framework.
- Geodesic convexity explains why simple greedy algorithms can work.
- Made quantitative by NC generalization of convex programming duality.
- We provide two general algorithms for geodesic convex optimization (which solve problems in poly time for many interesting actions).

Framework: Noncommutative group optimization [BFGOWW]

Action of "nice" $G \subseteq \mathrm{GL}_n$ on $V \cong \mathbb{C}^m$.



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

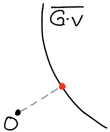
Norm Minimization | Scaling Problem

- All examples mentioned earlier fall into this framework.
- Geodesic convexity explains why simple greedy algorithms can work.
- Made quantitative by NC generalization of convex programming duality.
- We provide two general algorithms for geodesic convex optimization (which solve problems in poly time for many interesting actions).

Framework: Noncommutative group optimization [BFGOWW]

Action of "nice" $G \subseteq \mathrm{GL}_n$ on $V \cong \mathbb{C}^m$.

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

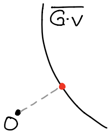
Norm Minimization — Scaling Problem

- All examples mentioned earlier fall into this framework.
- Geodesic convexity explains why simple greedy algorithms can work.
- Made quantitative by NC generalization of convex programming duality.
- We provide two general algorithms for geodesic convex optimization (which solve problems in poly time for many interesting actions).

Framework: Noncommutative group optimization [BFGOWW]

Action of "nice" $G \subseteq \mathrm{GL}_n$ on $V \cong \mathbb{C}^m$.

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

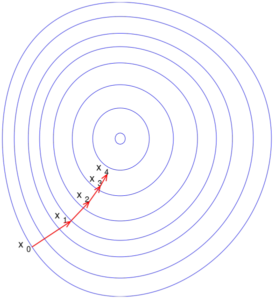


Minimize $\|g \cdot v\|$ over $g \in G$. $\underset{\text{NC-duality}}{\overset{\text{g-convexity}}{\Longleftrightarrow}}$ Find $g \in G$ s.th. $\nabla \|g \cdot v\| \approx 0$.

Norm Minimization — Scaling Problem

- All examples mentioned earlier fall into this framework.
- Geodesic convexity explains why simple greedy algorithms can work.
- Made quantitative by NC generalization of convex programming duality.
- We provide two general algorithms for geodesic convex optimization (which solve problems in poly time for many interesting actions).

Algorithms



First order algorithm for scaling ("gradient descent")

Idea: Repeatedly perform geodesic gradient steps

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

Theorem

Let $v \in V$ be not in the null cone. Then the algorithm outputs $g \in G$ such that $\|\nabla F(g)\| \leqslant \varepsilon$ 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

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

Theorem

Let $v \in V$ be not in the null cone. Then the algorithm outputs $g \in G$ such that $\|\nabla F(g)\| \leqslant \varepsilon$ 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

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

Theorem

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

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 smoothness L, latter by $\frac{1}{\gamma}$.

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

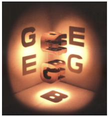
Let $v \in V$ be not in the null cone. Then the algorithm outputs $g \in G$ such that $F(g) \leqslant \inf_{g \in G} F(g) + \varepsilon$ within $T = \text{poly}(\log \frac{1}{\varepsilon}, \text{input size}, \frac{1}{\gamma})$ steps.

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

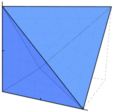
State of the art: Two general algorithms for geodesic convex optimization, which can solve norm minimization, scaling, null cone. Polynomial time for many interesting actions – but not always!

Summary and outlook

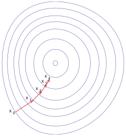
Symmetries lie behind many natural computational problems from algebra and analysis to classical and quantum CS.



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



Symmetries are key to tackling problems by optimization. Enabled by geodesic convexity and invariant theory.



Many exciting open questions: Poly-time algorithms for general actions? Better tools for geodesic convex optimization? Structured or typical data? Other problems with natural symmetries? **Thank you for your attention!**