Tensors, invariants, and optimization

Michael Walter

Perimeter Institute, June 2020

based on joint works with Peter Bürgisser, Cole Franks, Ankit Garg, Rafael Oliveira, Avi Wigderson (ITCS’18, FOCS’18, FOCS’19)
Overview

There are geometric and algebraic problems, originating in invariant theory, that are amenable to numerical optimization algorithms over groups.

Marginal & scaling problems $\leftrightarrow$ Null cone problems

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. Algorithmic solution

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 algorithms are known for graph isomorphism
- matrices equivalent under row and column operations iff equal rank; but tensor rank is NP-hard
- derandomizing PIT implies circuit lower bounds [Kabanets-Impagliazzo]
- 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*.

![Symmetric Group Action Example]

**Problem:** How can we algorithmically and efficiently check equivalence?

Interesting (and often difficult) problems with many applications:

- no polynomial-time algorithms are known for *graph isomorphism*
- matrices equivalent under row and column operations iff equal rank; but *tensor rank* is NP-hard
- derandomizing PIT 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*.

![Symmetry diagram](image)

**Problem:** How can we algorithmically and efficiently check equivalence?

Interesting (and often difficult) problems with many applications:

- no polynomial-time algorithms are known for *graph isomorphism*
- matrices equivalent under row and column operations iff equal rank; but *tensor rank* is NP-hard
- derandomizing *PIT* implies circuit lower bounds
  
  [Kabanets-Impagliazzo]
- 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 algorithms are known for graph *isomorphism*
- matrices equivalent under row and column operations iff equal rank; but tensor *rank* is NP-hard
- derandomizing *PIT* implies circuit lower bounds
- computing *normal forms*, describing *moduli spaces* and *invariants*... 

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

**Group** $G \subseteq \text{GL}_n(\mathbb{C})$, such as $\text{GL}_n$, $\text{SL}_n$, or $T_n = (\mathbb{C}^*)^n$

**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: $v \in \overline{Gw}$? $\overline{Gv} \cap \overline{Gw} \neq \emptyset$?
- Null cone problem: $0 \in \overline{Gv}$?

Classical problems. The last two can be solved via invariants. Are there more efficient ways?
Setup and orbit problems

**Group** $G \subseteq \text{GL}_n(\mathbb{C})$, such as $\text{GL}_n$, $\text{SL}_n$, or $T_n = (\mathbb{C}^*)^n$

**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: $v \in \overline{Gw}$? $Gv \cap \overline{Gw} \neq \emptyset$?
- Null cone problem: $0 \in \overline{Gv}$?

Classical problems. The last two can be solved via invariants. Are there more efficient ways?
Setup and orbit problems

**Group** $G \subseteq \text{GL}_n(\mathbb{C})$, such as $\text{GL}_n$, $\text{SL}_n$, or $T_n = (\mathbb{C}^*)^n$

**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: $v \in \overline{Gw}$? $\overline{Gv} \cap \overline{Gw} \neq \emptyset$?
- Null cone problem: $0 \in \overline{Gv}$?

Classical problems. The last two can be solved via invariants. Are there more efficient ways?
Setup and orbit problems

**Group** \( G \subseteq \text{GL}_n(\mathbb{C}) \), such as \( \text{GL}_n \), \( \text{SL}_n \), or \( T_n = (\mathbb{C}^*)^n \)

**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: \( v \in \overline{Gw} \)? \( \overline{Gv} \cap \overline{Gw} \neq \emptyset \)?
- Null cone problem: \( 0 \in \overline{Gv} \)?

Classical problems. The last two can be solved via invariants. Are there more efficient ways?
Example: Conjugation

\[ G = GL_n, \quad V = \text{Mat}_n, \quad g \cdot X = gXg^{-1} \]

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

- \( X, Y \) 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

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

For concreteness, focus on the null cone problem: Is $0 \in Gv$?

We can translate this into an optimization problem on the group $G$:

$$\inf_{g \in G} \| g \cdot v \| = ?$$

First-order condition? Clearly, the gradient at any minimizer $g$ is zero. Remarkably, this is also sufficient!

Thus, we can equivalently minimize the gradient. Moreover, in many applications the gradient is object of primary interest!
Orbit problems and optimization

For concreteness, focus on the null cone problem: Is $0 \in \overline{Gv}$?

We can translate this into an optimization problem on the group $G$:

$$\inf_{g \in G} \|g \cdot v\| = ?$$

**First-order condition?** Clearly, the gradient at any minimizer $g$ is zero. Remarkably, this is also sufficient!

Thus, we can equivalently minimize the gradient. Moreover, in many applications the gradient is object of primary interest!
Orbit problems and optimization

For concreteness, focus on the null cone problem: Is $0 \in \overline{Gv}$?

We can translate this into an optimization problem on the group $G$:

$$\inf_{g \in G} \|g \cdot v\| = ?$$

First-order condition? Clearly, the gradient at any minimizer $g$ is zero. Remarkably, this is also sufficient!

Thus, we can equivalently minimize the gradient. Moreover, in many applications the gradient is object of primary interest!
Summary so far

\[ G \subseteq \text{GL}_n \text{ acting linearly on } V = \mathbb{C}^m \]

**Null cone problem:** Given \( v \), is \( 0 \in \overline{Gv} \)?)

...and its relaxations:

**Norm minimization problem:** Given \( v \), find \( g \in G \) s.th. \( \| g \cdot v \| \approx \inf \).

**Scaling problem:** Given \( v \in V \), find \( g \in G \) s.th. \( \nabla \| g \cdot v \| \approx 0 \).

▶ The last two problems are dual, and either can solve null cone!
▶ But they also provide path to other orbit problems.

Useful *model problems*. Plausibly solvable in polynomial time, but rich enough to have interesting applications. Let us look at some...
Summary so far

\[ G \subseteq \text{GL}_n \text{ acting linearly on } V = \mathbb{C}^m \]

**Null cone problem:** Given \( \nu \), is \( 0 \in \overline{G\nu} \)?

...and its relaxations:

**Norm minimization problem:** Given \( \nu \), find \( g \in G \) s.t. \( \|g \cdot \nu\| \approx \inf \).

**Scaling problem:** Given \( \nu \in V \), find \( g \in G \) s.t. \( \nabla \|g \cdot \nu\| \approx 0 \).

- The last two problems are dual, and either can solve null cone!
- But they also provide path to other orbit problems.

Useful *model problems*. Plausibly solvable in polynomial time, but rich enough to have interesting applications. Let us look at some...
A panorama of applications
Example: Matrix scaling (raking, IPFP, . . .)

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

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

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

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

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

Permanent (Algebra): . . . iff $\text{per}(X) > 0!$

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

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

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

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

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

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

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

- . . . iff ∃ bipartite perfect matching in support of $X$
- can be decided in polynomial time
- find scalings by alternatingly fixing rows & columns 😊

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

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

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

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

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

Permanent (Algebra): . . . iff \( \text{per}(X) > 0! \)

$\Rightarrow$ iff $\exists$ bipartite perfect matching in support of $X$

$\Rightarrow$ can be decided in polynomial time

$\Rightarrow$ find scalings by alternatingly fixing rows & columns

Sinkhorn

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

\[ V = \text{Mat}_n, \quad G = T_n \times T_n, \quad (g_1, g_2) v = g_1 v g_2. \]

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

\[ V = \text{Mat}_n, \quad G = T_n \times T_n, \quad (g_1, g_2) v = g_1 v g_2. \]
Example: Sinkhorn algorithm

\[
\begin{pmatrix}
1 & 2 \\
4 & 0
\end{pmatrix}
\xrightarrow{\text{fix rows}}
\begin{pmatrix}
\frac{1}{3} & \frac{2}{3} \\
1 & 0
\end{pmatrix}
\xrightarrow{\text{fix cols}}
\begin{pmatrix}
\frac{1}{4} & 1 \\
\frac{3}{4} & 0
\end{pmatrix}
\rightarrow \cdots \rightarrow
\begin{pmatrix}
\frac{1}{2t-1} & 1 \\
\frac{2t}{2t} & 0
\end{pmatrix}
\]

after \( t \) steps. Why does it work? Permanent increases monotonically – can be used to control convergence:

- permanent
- distance to doubly stochastic

State-of-the-art algorithms directly optimize the norm square (in disguise).
Example: Sinkhorn algorithm

\[
\begin{pmatrix}
1 & 2 \\
4 & 0
\end{pmatrix}
\text{ fix rows } \rightarrow
\begin{pmatrix}
\frac{1}{3} & \frac{2}{3} \\
1 & 0
\end{pmatrix}
\text{ fix cols } \rightarrow
\begin{pmatrix}
\frac{1}{3} & 1 \\
\frac{4}{3} & 0
\end{pmatrix}
\rightarrow \ldots \rightarrow
\begin{pmatrix}
\frac{1}{2t} & 1 \\
\frac{2t-1}{2t} & 0
\end{pmatrix}
\]

after \(t\) steps. Why does it work? Permanent increases monotonically – can be used to control convergence:

<table>
<thead>
<tr>
<th>(50)</th>
<th>(100)</th>
<th>(150)</th>
<th>(200)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.236</td>
<td>0.238</td>
<td>0.240</td>
<td>0.242</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(50)</th>
<th>(100)</th>
<th>(150)</th>
<th>(200)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0005</td>
<td>0.0010</td>
<td>0.0015</td>
<td>0.0020</td>
</tr>
</tbody>
</table>

State-of-the-art algorithms directly optimize the norm square (in disguise).
Example: Operator scaling and non-commutative PIT

Let $T(\rho) = \sum_i X_i \rho X_i^\dagger$ be a CP map. A *scaling* of $T$ is of the form

$$S(\rho) = AT(B\rho B^\dagger)A^\dagger \quad (A, B \in \text{GL}_n)$$

Say $T$ is *quantum doubly stochastic* if $T(I) = T^\dagger(I) = I$.

**Operator scaling:** Given $T$, $\exists$ approximately quantum d.s. scalings?

**Polynomial identity testing:** ... iff $\exists$ matrices $Y_k$ s.th. $\det \sum_k Y_k \otimes X_k \neq 0$.

- natural iterative algorithm: alternatingly make unital and trace-preserving [Gurvits]
- can solve in deterministic polynomial time [Garg et al, Ivanyos et al]

When $Y_k$ restricted to scalars? Major open problem in TCS!
Example: Operator scaling and non-commutative PIT

Let $T(\rho) = \sum_i X_i \rho X_i^\dagger$ be a CP map. A scaling of $T$ is of the form

$$S(\rho) = AT(B\rho B^\dagger)A^\dagger$$

$(A, B \in \text{GL}_n)$

Say $T$ is quantum doubly stochastic if $T(I) = T^\dagger(I) = I$.

**Operator scaling:** Given $T$, $\exists$ approximately quantum d.s. scalings?

**Polynomial identity testing:** \ldots iff $\exists$ matrices $Y_k$ s.th. $\det \sum_k Y_k \otimes X_k \neq 0$.

- natural iterative algorithm: alternatingly make unital and trace-preserving [Gurvits]
- can solve in deterministic polynomial time [Garg et al, Ivanyos et al]

When $Y_k$ restricted to scalars? Major open problem in TCS!
Applications and connections

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’s M-approximation

Optimization: Efficient algorithms for classes of quadratic equations

Computational complexity: Polynomial identity testing, tensor ranks

Quantum information: Marginal problems, entanglement transformations

All these are special cases of a general class of problems! We now focus on one scenario that is in many ways ‘representative’.
Quantum states and marginals

Pure quantum state of $d$ particles is described by unit-norm tensor:

$$X \in V = \mathbb{C}^{n_1} \otimes \cdots \otimes \mathbb{C}^{n_d}$$

State of individual particles described by density matrices $\rho_1, \ldots, \rho_d$:

$$\text{tr}[\rho_1 H_1] = \langle X| H_1 \otimes I \otimes \cdots \otimes I |X \rangle \quad \forall H_1$$

Quantum marginal problem: Which $\rho_1, \ldots, \rho_d$ are consistent with a global pure state $X$?

Answer only depends on the eigenvalues $\lambda_i$ of $\rho_i$!
Quantum states and marginals

Pure quantum state of $d$ particles is described by unit-norm tensor:

$$X \in V = \mathbb{C}^{n_1} \otimes \ldots \otimes \mathbb{C}^{n_d}$$

State of individual particles described by density matrices $\rho_1, \ldots, \rho_d$:

$$\text{tr}[\rho_1 H_1] = \langle X| H_1 \otimes I \otimes \ldots \otimes I |X\rangle \quad \forall H_1$$

Quantum marginal problem: Which $\rho_1, \ldots, \rho_d$ are consistent with a global pure state $X$?

Answer only depends on the eigenvalues $\lambda_i$ of $\rho_i$!
Examples

Two particles: $\rho_A$ and $\rho_B$ compatible with global pure state iff same nonzero eigenvalues (Schmidt decomposition)

Three particles:

\[
\begin{align*}
\lambda_{A,max} + \lambda_{B,max} & \leq \lambda_{C,max} + 1 \\
\lambda_{A,max} + \lambda_{C,max} & \leq \lambda_{B,max} + 1 \\
\lambda_{B,max} + \lambda_{C,max} & \leq \lambda_{A,max} + 1
\end{align*}
\]

- necessary and sufficient for qubits
- follows from variational principle: $\lambda_{A,max} = \max_{\Phi_A} \langle \Phi_A | \rho_A | \Phi_A \rangle$ etc.

[Higuchi, Sudbery, Szulc]
Tensor scaling and SLOCC

A scaling of $X$ is a tensor of the form

$$Y = (A_1 \otimes \ldots \otimes A_d)X \quad (A_i \in \text{GL}_{n_i})$$

- state that can be obtained by SLOCC (postselected local operations & classical communication)
- $X$ constrains the entanglement class

Tensor scaling problem: Which $\rho_1, \ldots, \rho_d$ arise from scaling of given $X$?

- e.g. for $\rho_i \propto I$, each system maximally entangled with rest
  (= locally maximally mixed = quantum version of stochastic tensor)
- again, answer only depends on eigenvalues $\lambda_i$ of $\rho_i$
Tensor scaling and SLOCC

A *scaling* of $X$ is a tensor of the form

$$Y = (A_1 \otimes \ldots \otimes A_d)X \quad (A_i \in \text{GL}_{n_i})$$

- state that can be obtained by SLOCC (postselected local operations & classical communication)
- $X$ constrains the entanglement class

**Tensor scaling problem:** Which $\rho_1, \ldots, \rho_d$ arise from scaling of given $X$?

- e.g. for $\rho_i \propto I$, each system *maximally entangled* with rest
  ($= \text{locally maximally mixed} = \text{quantum version of stochastic tensor}$)
- again, answer only depends on eigenvalues $\lambda_i$ of $\rho_i$
Tensor scaling and entanglement polytopes

Thus, answer to tensor scaling problem is encoded by:

\[ \Delta(X) = \left\{ (\lambda_1, \ldots, \lambda_d) \mid \text{for scalings of } X \text{ (and limits)} \right\} \subseteq \mathbb{R}^{dn} \]

e.g., for three qubits, \( \text{GHZ} = |000\rangle + |111\rangle \) and \( \text{W} = |100\rangle + |010\rangle + |001\rangle \):

In general, always convex polytopes:
- encode local info about entanglement
- encode recent notions of tensor ranks

However, explicit description intractable. Exponential number of vertices and facets!

We provide algorithmic solution!
Tensor scaling and entanglement polytopes

Thus, answer to tensor scaling problem is encoded by:

\[ \Delta(X) = \{(\lambda_1, \ldots, \lambda_d) \text{ for scalings of } X \text{ (and limits)} \} \subseteq \mathbb{R}^{dn} \]

e.g., for three qubits, \( \text{GHZ} = |000\rangle + |111\rangle \) and \( \text{W} = |100\rangle + |010\rangle + |001\rangle \):

In general, always convex polytopes:

- encode local info about entanglement
- encode recent notions of tensor ranks

However, explicit description intractable. Exponential number of vertices and facets!

We provide algorithmic solution!
Tensor scaling and entanglement polytopes

Thus, answer to tensor scaling problem is encoded by:

\[ \Delta(X) = \left\{ (\lambda_1, \ldots, \lambda_d) \text{ for scalings of } X \text{ (and limits)} \right\} \subseteq \mathbb{R}^{dn} \]

E.g., for three qubits, \( \text{GHZ} = |000\rangle + |111\rangle \) and \( \text{W} = |100\rangle + |010\rangle + |001\rangle \):

In general, always convex polytopes:

- encode local info about entanglement
- encode recent notions of tensor ranks

However, explicit description intractable. Exponential number of vertices and facets!

We provide algorithmic solution!
Tensor scaling and entanglement polytopes

Thus, answer to tensor scaling problem is encoded by:

\[ \Delta(X) = \left\{ (\lambda_1, \ldots, \lambda_d) \text{ for scalings of } X \text{ (and limits)} \right\} \subseteq \mathbb{R}^{dn} \]

\[ e.g., \text{ for three qubits, } GHZ = |000\rangle + |111\rangle \text{ and } W = |100\rangle + |010\rangle + |001\rangle: \]

In general, always convex polytopes:

- encode local info about entanglement
- encode recent notions of tensor ranks

However, explicit description intractable. Exponential number of vertices and facets!

\[ \text{We provide algorithmic solution!} \]
Geodesic optimization algorithms
The Algorithm

Given state $X$, want to find scaling $Y$ with desired marginals – whenever possible. For simplicity, uniform marginals $(\rho_i \propto 1, \lambda_i \propto 1)$ and $d = 3$.

**Algorithm:** Start with $Y = X$. For $t = 1, \ldots, T$:
Compute marginals $\rho_1, \rho_2, \rho_3$ of $Y$. If $\varepsilon$-close to uniform, stop.
Otherwise, replace $Y$ by $(e^{-\eta \rho_1^o} \otimes e^{-\eta \rho_2^o} \otimes e^{-\eta \rho_3^o})Y$. $X^o = \text{traceless part}$

$\eta = \text{suitable step size}$

**Theorem**
Algorithm finds $Y = (A_1 \otimes A_2 \otimes A_3)X$ with marginals $\varepsilon$-close to uniform within $T = \text{poly}(\frac{1}{\varepsilon}, \text{input size})$ steps.

- generalizes to arbitrary $\lambda_i$, $d > 3$, (anti)symmetric tensors, MPS, …
- can run on quantum computer (but how well? 😊)
- solve quantum marginal problem by using random $X$

cf. algorithm by Verstraete et al which we analyzed in prior work
Consider the problem of minimizing the norm

\[ N(A_1, A_2, A_3) = \| (A_1 \otimes A_2 \otimes A_3)X \| \quad (A_i \in \text{SL}_{n_i}) \]

Its derivative in direction given by traceless \( H_1, H_2, H_3 \) is

\[ \partial_{t=0} N(e^{tH_1}, e^{tH_2}, e^{tH_3}) = \text{tr}[\rho_1^o H_1] + \text{tr}[\rho_2^o H_2] + \text{tr}[\rho_3^o H_3]. \]

Therefore, the gradient can be identified with \( \nabla N = (\rho_1^o, \rho_2^o, \rho_3^o) \).

- Algorithm implements geodesic gradient descent...
- ...and minimizing the gradient makes the marginals uniform! 😊

How to make quantitative? What is the big picture?
Consider the problem of minimizing the norm
\[ N(A_1, A_2, A_3) = \| (A_1 \otimes A_2 \otimes A_3) X \| \quad (A_i \in \text{SL}_{n_i}) \]
Its derivative in direction given by traceless \( H_1, H_2, H_3 \) is
\[ \partial_{t=0} N(e^{tH_1}, e^{tH_2}, e^{tH_3}) = \text{tr}[\rho_1^o H_1] + \text{tr}[\rho_2^o H_2] + \text{tr}[\rho_C^o H_3]. \]
Therefore, the gradient can be identified with \( \nabla N = (\rho_1^o, \rho_2^o, \rho_3^o) \).

▶ Algorithm implements geodesic gradient descent
▶ ...and minimizing the gradient makes the marginals uniform!

How to make quantitative? What is the big picture?
Why does it work?

“Otherwise, replace $Y$ by $(e^{-\eta \rho_1} \otimes e^{-\eta \rho_2} \otimes e^{-\eta \rho_3}) Y$.”

Consider the problem of minimizing the norm

$$N(A_1, A_2, A_3) = \|(A_1 \otimes A_2 \otimes A_3) X\| \quad (A_i \in \text{SL}_{n_i})$$

Its derivative in direction given by traceless $H_1, H_2, H_3$ is

$$\partial_{t=0} N(e^{tH_1}, e^{tH_2}, e^{tH_3}) = \text{tr}[\rho_1 H_1] + \text{tr}[\rho_2 H_2] + \text{tr}[\rho_3 H_3].$$

Therefore, the gradient can be identified with $\nabla N = (\rho_1, \rho_2, \rho_3)$.

- Algorithm implements geodesic gradient descent...
- ...and minimizing the gradient makes the marginals uniform! 😊

How to make quantitative? What is the big picture?
Non-commutative optimization

In general, consider $N(g) = \|g \cdot X\|$.

We discussed that the following optimization problems are equivalent:

$$\inf_{g \in G} N(g) \iff \inf_{g \in G} \|\nabla N(g)\|$$

- **primal**: norm minimization, **dual**: scaling problem
- non-commutative version of linear programming duality

We develop quantitative duality theory and 1st & 2nd order methods.

All examples from introduction fall into this framework.

Numerical algorithms that solve algebraic problems!

Everything works for general actions of reductive $G$. Norm is log-convex along geodesics.
Geodesic convexity

Why does the duality hold? Consider geodesics $g_t = e^{tH}g$ in the group $G$.

![Geodesic](image)

**Proposition:** $N(g) = \|g \cdot \nu\|$ satisfies along these geodesics:

1. **convexity:** $\partial^2_{t=0} N(g_t) \geq 0$
2. **smoothness:** $\partial^2_{t=0} N(g_t) \leq 2C^2 \|H\|_F^2$

$C$ is typically small, upper-bounded by degree of action.

**Smoothness** implies that

$$N(e^{H}g) \leq N(g) + \nabla N(g) \cdot H + C^2 \|H\|_F^2.$$

Thus, gradient descent makes progress if steps not too large!
Analysis of Algorithm

“Unless $\varepsilon$-close to uniform, replace $Y$ by $(e^{-\eta \rho^1_1} \otimes e^{-\eta \rho^2_2} \otimes e^{-\eta \rho^3_3})Y$.”

To obtain rigorous algorithm, show:

- **progress in each step:** $\|Y_{\text{new}}\| \leq (1 - c_1 \varepsilon)\|Y\|$

- **a priori lower bound:** $\inf_{\det = 1}\|(A_1 \otimes A_2 \otimes A_3)X\| \geq c_2$

Then, $(1 - c_1 \varepsilon)^T \geq c_2$ bounds the number of steps $T$.

The first point follows from smoothness, as just discussed.

For the second, construct ‘explicit’ invariants with ‘small’ coefficients, so that $P(X) \neq 0$ implies bound in terms of bitsize of $X$. 
Analysis of Algorithm

“Unless $\varepsilon$-close to uniform, replace $Y$ by $(e^{-\eta \rho_1} \otimes e^{-\eta \rho_2} \otimes e^{-\eta \rho_3}) Y$.”

To obtain rigorous algorithm, show:

- **progress in each step:** $\| Y_{\text{new}} \| \leq (1 - c_1 \varepsilon) \| Y \|$
- **a priori lower bound:** $\inf_{\det = 1} \|(A_1 \otimes A_2 \otimes A_3) X\| \geq c_2$

Then, $(1 - c_1 \varepsilon)^T \geq c_2$ bounds the number of steps $T$.

The first point follows from smoothness, as just discussed.

For the second, construct ‘explicit’ invariants with ‘small’ coefficients, so that $P(X) \neq 0$ implies bound in terms of bitsize of $X$. 
Effective algorithms for large class of optimization problems over groups, incl. quantum marginal and tensor scaling problems. Based on geodesic convex optimization and geometric invariant theory.

Many exciting directions:

▶ Polynomial-time algorithms in all cases?
▶ Better tools for geodesic optimization? Quantum algorithms?
▶ Tensors in quantum information are often special. Implications?
▶ Can we tackle other problems with natural symmetries?

Thank you for your attention!
A general equivalence \( \mathcal{V} \subseteq \mathbb{P}(\mathcal{V}) \)

All points in \( \Delta(\mathcal{V}) \) can be described via invariant theory:

\[
\mathcal{V}_\lambda \subseteq \mathbb{C}[\mathcal{V}]_{(k)} \implies \frac{\lambda}{k} \in \Delta(\mathcal{V})
\]

(\( \lambda \) highest weight, \( k \) degree)

- Can also study multiplicities \( g(\lambda, k) := \# \mathcal{V}_\lambda \subseteq \mathbb{C}[\mathcal{V}]_{(k)} \).

- This leads to interesting computational problems:

  \( g = ? \)

  \( g > 0 ? \)

  \( \exists s > 0 : g(s\lambda, sk) > 0 ? \)

  (\( \# \)-hard)

  (NP-hard)

  (our problem!)

Completely unlike Horn’s problem: *Knutson-Tao saturation property does not hold, and hence we can hope for efficient algorithms!*