Title: Tensors, invariants, and optimization

Speakers: Michael Walter

Series: Perimeter Institute Quantum Discussions

Date: June 17, 2020 - 4:00 PM

URL: http://pirsa.org/20060049

Abstract: Given a vector in a representation, can it be distinguished from zero by an invariant polynomial? This classical question in invariant theory relates to a diverse set of problems in mathematics and computer science. In quantum information, it captures the quantum marginal problem and recent bounds on tensor ranks. We will see that the general question can be usefully thought of as an optimization problem and discuss how this perspective leads to efficient algorithms for solving it.

Tensors, invariants, and optimization

Michael Walter

UNIVERSITY OF AMSTERDAM NWO

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.

$$\boxed{\text{Marginal \& scaling problems}} \longleftrightarrow \boxed{\text{Null cone problems}}$$

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

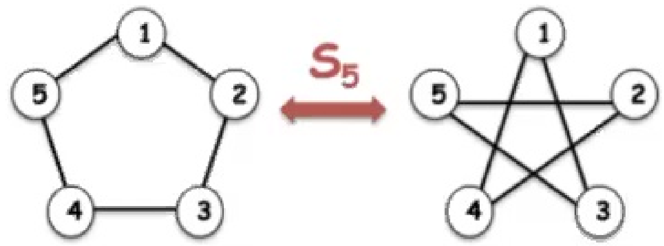
Plan for today:

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. . .

Setup and orbit problems

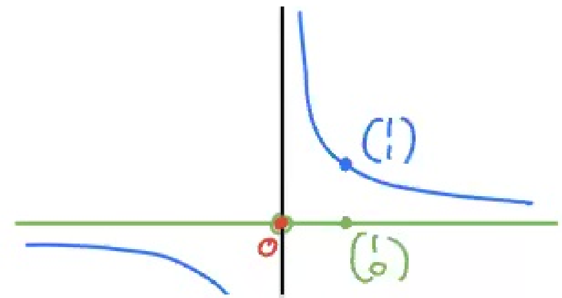
Group $G \subseteq \mathrm{GL}_n(\mathbb{C})$, such as GL_n, SL_n, or $\mathrm{T}_n = (\mathbb{C}^*)^n$

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

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

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

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



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

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

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

- X, Y are in *same orbit* iff same Jordan normal form
- X, Y have *intersecting orbit closures* iff same eigenvalues
- X is in *null cone* iff nilpotent

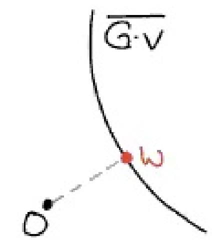
NB: The last two problems have a meaningful approximate version!

Orbit problems and optimization

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

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

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

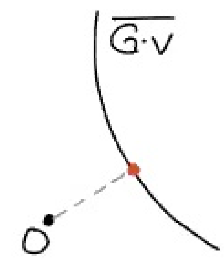


First-order condition? Clearly, the gradient at any minimizer g is zero.
Remarkably, this is also sufficient! [Kempf-Ness]

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

Summary so far

$G \subseteq \mathrm{GL}_n$ acting linearly on $V = \mathbb{C}^m$



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

...and its relaxations:

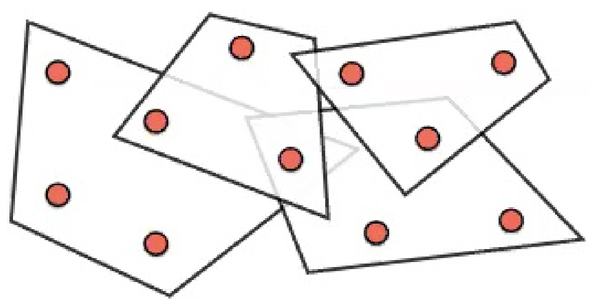
Norm minimization problem: Given v, find $g \in G$ s.th. $\|g \cdot v\| \approx \mathrm{cap}(v)$.

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...

A panorama of applications



Example: Matrix scaling (raking, IPFP, . . .)

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

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

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

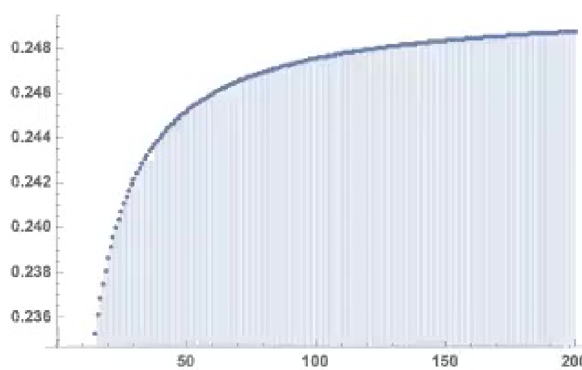
- . . . iff $\exists$ bipartite perfect matching in support of X
- can be decided in polynomial time
- find scalings by alternatingly fixing rows & columns ☺ [Sinkhorn]

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

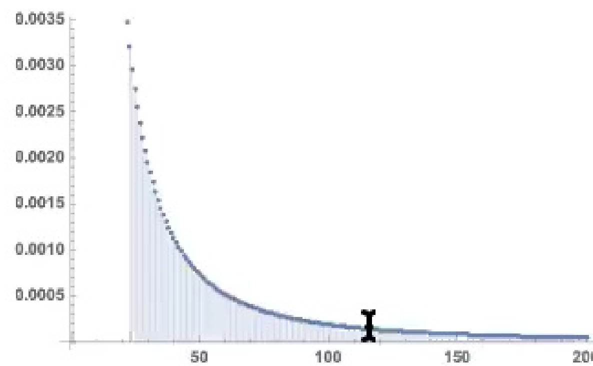
Example: Sinkhorn algorithm

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

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



permanent



distance to doubly stochastic

State-of-the-art algorithms directly optimize the norm square (in disguise).

Example: Matrix scaling (raking, IPFP, ...)

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

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

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

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

Pe

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

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

orn]

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

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

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

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!

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 $\quad \exists\ A + B = C$ with spectrum α, β, γ?
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

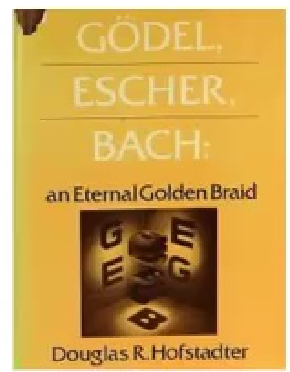
Pure quantum state of d particles is described by unit-norm tensor:

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

$X \;\boxed{1 \quad 2 \quad \cdots \quad d}$

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

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

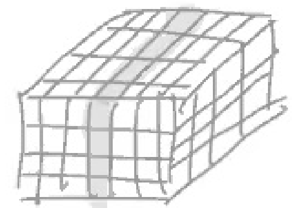


> Quantum marginal problem: Which $\rho_1, \dots, \rho_d$ are consistent with a global pure state X?

Answer only depends on the eigenvalues λ_i of ρ_i!

Tensor scaling and SLOCC

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



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

- 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?

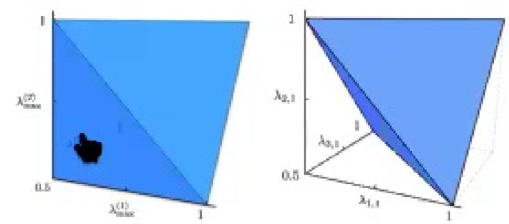
- e.g. for $\rho_i \propto I$, each system maximally entangled with rest
  (= locally maximally mixed = quantum version of stochastic tensor)
- again, answer only depends on eigenvalues λ_i of ρ_i

Tensor scaling and entanglement polytopes

Thus, answer to tensor scaling problem is encoded by:

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

e.g., for three qubits, $GHZ = |000\rangle + |111\rangle$ *and* $W = |100\rangle + |010\rangle + |001\rangle$*:*



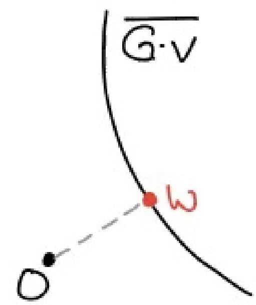
In general, always convex polytopes: [Kirwan, Mumford]

- encode local info about entanglement [W-Christandl-Doran-Gross, Sawicki et al]
- encode recent notions of tensor ranks [Christandl et al, Derksen]

However, explicit description intractable. [Berenstein-Sjamaar, Klyachko, Ressayre, Vergne-W.]
Exponential number of vertices and facets!

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 ($\rho_i \propto I$, $\boldsymbol{\lambda}_i \propto 1$) and $d = 3$.

Algorithm: Start with $Y = X$. For $t = 1, \dots, T$:
Compute marginals ρ_1, ρ_2, ρ_3 of Y. If ε-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 = *traceless part*

η = *suitable step size*

Theorem

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

- generalizes to arbitrary $\boldsymbol{\lambda}_i$, $d > 3$, (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

Why does it work?

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

Consider the problem of minimizing the norm

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

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

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

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?

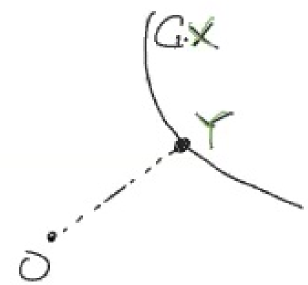
Non-commutative optimization

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

We discussed that the following *optimization problems* are equivalent:

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

[Kempf-Ness]



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

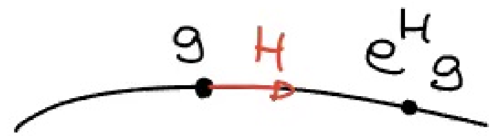
We develop quantitative duality theory and 1st & 2nd order methods.

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.



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

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

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

Smoothness implies that

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

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

Analysis of Algorithm

"Unless ε-close to uniform, replace Y by $(e^{-\eta\rho_1^o} \otimes e^{-\eta\rho_2^o} \otimes e^{-\eta\rho_3^o})Y$."

To obtain rigorous algorithm, show:

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

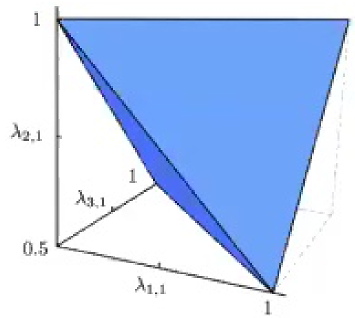
Then, $(1 - c_1\varepsilon)^T \geqslant c_2$ bounds the number of steps T.

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.

gradflow_waterloo.pdf (page 44 of 45)

Summary and outlook



Marginal & scaling problems

$\updownarrow$ duality

Norm minimization & null cone

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!