Tensor scaling, quantum marginals, and moment polytopes

Michael Walter

University of Amsterdam QuSoft Research Center for Quantum Software NWO

SIAM Conference on Applied Algebraic Geometry, Bern, July 2019

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

Overview: Scaling and marginal problems

Interesting class of problems — with applications in q. information, algebra, analysis, computer science — that *surprisingly* can be phrased as *optimization problems* over noncommutative groups.

Null cone & moment polytopes	$\longleftrightarrow$	Norm minimization
(Geometric invariant theory)		(Optimization theory)

Result: General framework and effective algorithms.

Plan: Overview and illustration via tensor scaling problem.

Example: Matrix scaling

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

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

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

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

Permanent (Invariant Theory): ...iff $\mathrm{per}(X) > 0$!

- can be decided in polynomial time
- find scalings by alternatingly fixing rows & columns ☺ [Sinkhorn]
- convergence controlled by permanent [Linial et al]

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

Example: Matrix scaling

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

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

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 (Invariant Theory): ...iff $\mathrm{per}(X) > 0$!

- can be decided in polynomial time
- find scalings by alternatingly fixing rows & columns ☺ [Sinkhorn]
- convergence controlled by permanent [Linial et al]

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

Further examples

- Horn problem $\exists$ Hermitian matrices $A + B = C$ with spectrum α, β, γ? [Franks]
- Positivity of Littlewood-Richardson coefficients [Knutson-Tao]

- Operator scaling [Gurvits, Garg et al, Ivanyos et al]
- Non-commutative polynomial identity testing

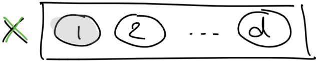
- Validity of Brascamp-Lieb inequalities [Bennett et al, Garg et al]
- Solution of Paulsen problem [Kwok et al]

All these are special cases of a general class of problems. Let us focus on 'representative' example involving tensors...

Quantum states and marginals

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

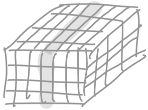
$$X \in V = (\mathbb{C}^n)^{\otimes d} = \mathbb{C}^n \otimes \cdots \otimes \mathbb{C}^n$$



State of individual particles described by quantum marginals $\rho_1,\ldots,\rho_d$:

$\rho_k = X_k X_k^*$, where X_k is k-th principal flattening of X

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



Answer only depends on eigenvalues λ_i of ρ_i!

(e.g., for $d = 2$: consistent iff same eigenvalues)

Quantum states and marginals

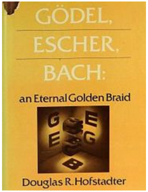
Global quantum state of d particles is described by unit-norm tensor

$$X \in V = (\mathbb{C}^n)^{\otimes d} = \mathbb{C}^n \otimes \cdots \otimes \mathbb{C}^n$$

State of individual particles described by quantum marginals $\rho_1,\ldots,\rho_d$:

$$\rho_k = X_k X_k^*, \text{ where } X_k \text{ is } k\text{-th principal flattening of } X$$

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

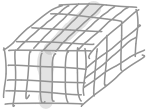


Answer only depends on eigenvalues λ_i of ρ_i!

(e.g., for $d = 2$: consistent iff same eigenvalues)

Tensor scaling and moment polytopes

Scaling of X: Tensor of the form $Y = (A_1 \otimes \ldots \otimes A_d) X$.



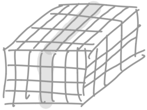
Tensor scaling problem: Given X, which $\lambda_1, \ldots, \lambda_d$ are consistent with its scalings (and limits)?

- $\{(\lambda_1, \ldots, \lambda_d)\}$ convex moment polytopes [Kirwan, Mumford]
- encode local info about entanglement [W-Christandl-Doran-Gross, Sawicki et al]
- exp. large V/H-representations [Berenstein-Sjamaar, Klyachko, Ressayre, Vergne-W]

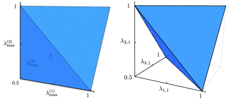
We provide algorithmic solution!

Tensor scaling and moment polytopes

Scaling of X: Tensor of the form $Y = (A_1 \otimes \ldots \otimes A_d) X$.



Tensor scaling problem: Given X, which $\lambda_1, \ldots, \lambda_d$ are consistent with its scalings (and limits)?

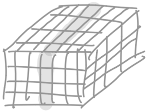


- $\{(\lambda_1, \ldots, \lambda_d)\}$ convex moment polytopes [Kirwan, Mumford]
- encode local info about entanglement [W-Christandl-Doran-Gross, Sawicki et al]
- exp. large V/H-representations [Berenstein-Sjamaar, Klyachko, Ressayre, Vergne-W]

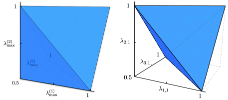
We provide algorithmic solution!

Tensor scaling and moment polytopes

Scaling of X: Tensor of the form $Y = (A_1 \otimes \ldots \otimes A_d)X$.



Tensor scaling problem: Given X, which $\lambda_1, \ldots, \lambda_d$ are consistent with its scalings (and limits)?



- $\{(\lambda_1, \ldots, \lambda_d)\}$ convex moment polytopes [Kirwan, Mumford]
- encode local info about entanglement [W-Christandl-Doran-Gross, Sawicki et al]
- exp. large V/H-representations [Berenstein-Sjamaar, Klyachko, Ressayre, Vergne-W]

We provide algorithmic solution!

An Algorithm

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

Algorithm: Start with $Y = X$. For $t = 1, \ldots, T$:
Compute marginals ρ_1, ρ_2, ρ_3 of Y. If ε-close to uniform, stop.
Otherwise, replace Y by $(e^{-\delta \rho_1^o} \otimes e^{-\delta \rho_2^o} \otimes e^{-\delta \rho_3^o}) Y$. *$X^o$ = traceless part*

Result

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

- generalizes to arbitrary λ_i, d>3, (anti)symmetric tensors, MPS, ...
- 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 $(e^{-\delta\rho_1^o} \otimes e^{-\delta\rho_2^o} \otimes e^{-\delta\rho_3^o})Y$."

This step implements *gradient descent* for logarithm of

$$N(A_1, A_2, A_3) = \|(A_1 \otimes A_2 \otimes A_3)X\|$$

where A_1,A_2,A_3 have det=1. Indeed:

- geodesic gradient can be identified with $(\rho_1^o, \rho_2^o, \rho_3^o)$!
- vanishes iff marginals uniform ☺

Why does it work?

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

This step implements *gradient descent* for logarithm of

$$N(A_1, A_2, A_3) = \|(A_1 \otimes A_2 \otimes A_3)X\|$$

where A_1,A_2,A_3 have det=1. Indeed:

- geodesic gradient can be identified with $(\rho_1^o, \rho_2^o, \rho_3^o)$!
- vanishes iff marginals uniform ☺

Non-commutative duality

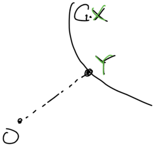
e.g. $G = \mathrm{SL}(n)^d$

For $N(g) = \|g \cdot X\|$, the following *optimization problems* are equivalent:

$$\boxed{\inf_{g \in G} N(g) > 0} \iff \boxed{\inf_{g \in G} \|\nabla \log N(g)\| = 0}$$

[Kempf-Ness]

moment map



- primal: norm minimization, dual: scaling problem
- non-commutative version of LP duality
- equivalent to semistability of X

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.

Analysis of Algorithm

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

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 geodesic convexity estimates.

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

To obtain rigorous algorithm, show:

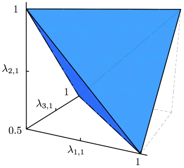
- *progress in each step:* $\|Y_{\mathrm{new}}\| \leq (1 - c_1\varepsilon)\|Y\|$
- *a priori lower bound:* $\inf_{\det=1}\|(A_1 \otimes A_2 \otimes A_3)X\| \geq c_2$

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

The first point follows from geodesic convexity estimates.

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

Summary and outlook



Marginal & scaling problems

$\updownarrow$ duality

Norm minimization

Effective algorithms for null cone and moment polytope problems, with applications incl. quantum marginal and tensor scaling problems. Based on geometric invariant theory and g-convex optimization.

Many exciting directions:

- Numerical studies in q. many-body systems or chemistry
- Quantum algorithms?
- Algorithms for other problems with natural symmetries?
- *What are the 'tractable' problems in invariant theory?* $\mathbb{C} \rightsquigarrow \mathbb{F}$?

Thank you for your attention!

A general equivalence

$\mathcal{V} \subseteq \mathbb{P}(V)$

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

$$V_\lambda \subseteq \mathbb{C}[\mathcal{V}]_{(k)} \quad \Rightarrow \quad \frac{\lambda}{k} \in \Delta(\mathcal{V})$$

(λ highest weight, k degree)

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