Hidden symmetries in computational problems (and geodesic convexity)

Avi Wigderson

IAS, Princeton

Zeyuan Allen-Zhu, Peter Burgisser, Cole Franks, Ankit Garg, Leonid Gurvits, Pavel Hrubes, Yuanzhi Li, Visu Makam, Rafael Oliveira, Michael Walter

Plan

One problem
Singularity of Symbolic Matrices
One algorithm
Alternating minimization

…internalize
…generalize (algorithms, problems, tools)

Extending convex optimization in Euclidean space to (geodesic) convex optimization on Riemannian manifolds, quantitative bounds

Applications & Connections

Non-commutative Algebra

Word problem in free skew fields

Invariant Theory

Symmetries, nullcone membership & orbit problems

Quantum Information Theory

Positive operators, quantum marginals

Analysis

Brascamp-Lieb inequalities

Operator Theory

Pauslen's problem on Parseval frames

Statistics

MLE in Gaussian models, Tyler's M-approximation

Computational complexity

Polynomial identity testing, arithmetic lower bounds

Optimization

Efficiently solving certain general families of

- Quadratic systems of equations
- Exponentially large linear LPs (Moment polytopes)

Optimization, Complexity and Math

through one problem and one algorithm

One problem

Singularity of Symbolic Matrices

One algorithm

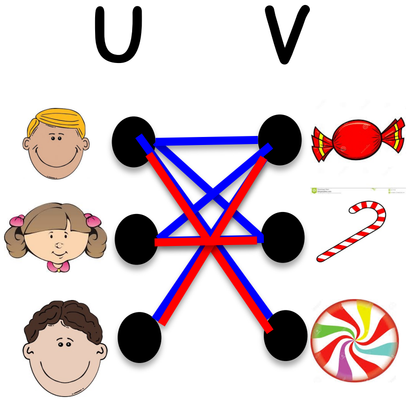
Alternating minimization

The problem(s)

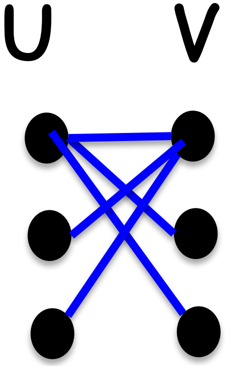
Perfect Matchings (PMs)

Bipartite graphs $G(U,V;E)$.

$|U|=|V|=n$



G'



G

$$\begin{array}{c|ccc} & & V & \\ \hline & 1 & 1 & 1 \\ U & 1 & 0 & 0 \\ & 1 & 0 & 0 \end{array}$$

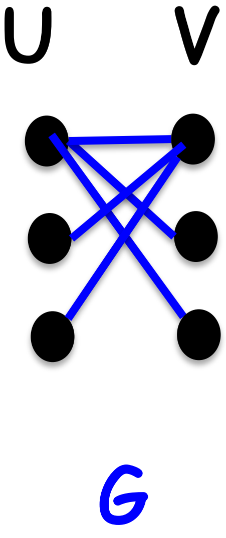
A_G

Fact: G has a PM iff $\mathrm{Per}(A_G)>0$

$$\mathrm{Per}_n(A)=\sum_{\sigma\in S_n}\prod_{i\in[n]} A_{i\sigma(i)}$$

[Jacobi'1890] PM $\in$ P (P = polynomial time)

PMs & symbolic matrices [Edmonds'67]



1	1	1
1	0	0
1	0	0

A_G

x_{11}	x_{12}	x_{13}
x_{21}	0	0
x_{31}	0	0

$A_G(X)$

[Edmonds '67] G has a PM iff $\text{Det}(A_G(X)) \neq 0$ $(\in P)$

Symbolic matrices [Edmonds'67]

$X = \{x_1, x_2, \ldots\}$ F field ($F=Q$)

$L_{ij}(X) = ax_1 + bx_2 + \ldots$:linear forms

$L(X) = A_1x_1 + A_2x_2 + \ldots + A_mx_m$ $A_i \in Mat_n(F)$

SING: Given ($A_1, \ldots A_m$) is $Det(L(X))=0$?

$$\begin{array}{|c|c|c|}
\hline
L_{11} & L_{12} & L_{13} \\
\hline
L_{21} & L_{22} & L_{23} \\
\hline
L_{31} & L_{32} & L_{33} \\
\hline
\end{array}$$

$L(X)$

[Edmonds '67] SING $\in$ P ??

[Lovasz '79] SING $\in$ RP

Randomized Poly Time

[Valiant '79] SING captures algebraic identities (PIT)

Math special cases: Module isomorphism, graph rigidity,…

[Kabanets-Impagliazzo'01] SING $\in$ P → "P ≠ NP"

Derandomization, Lower bounds

Symbolic matrices dual life

$X = \{x_1, x_2, \ldots x_m\}$ $\mathbb{F}$ field

$L(X) = A_1x_1 + A_2x_2 + \ldots + A_mx_m$

Input: $A_1, A_2, \ldots, A_m \in M_n(\mathbb{F})$

SING : Is $L(X)$ singular?

L_{11}	L_{12}	L_{13}
L_{21}	L_{22}	L_{23}
L_{31}	L_{32}	L_{33}

x_i commute

in $\mathbb{F}(x_1, x_2, \ldots, x_m)$

[Lovasz '79] SING $\in$ RP

[Edmonds'67] SING $\in$ P?

x_i do not commute

in $\mathbb{F}\langle(x_1, x_2, \ldots, x_m)\rangle$ (free skew field)

[Cohn'75] NC-SING Decidable

[CR'99] NC-SING $\in$ EXP

[GGOW'15] NC-SING $\in$ P ($\mathbb{F}=\mathbb{Q}$)

[IQS'16] NC-SING $\in$ P ($\mathbb{F}$ large)

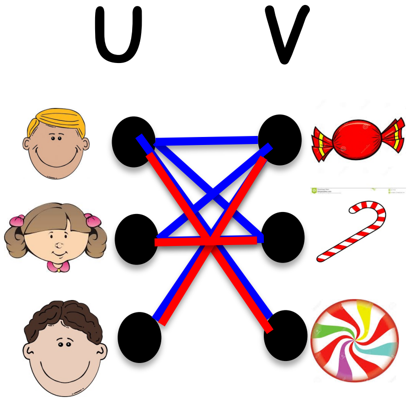
The algorithm

Alternate minimization

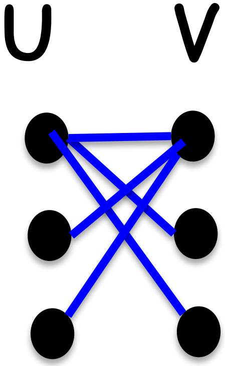
Perfect Matchings (PMs)

Bipartite graphs $G(U,V;E)$.

$|U|=|V|=n$



G'



G

$$\begin{array}{c|ccc} & & V & \\ \hline U & 1 & 1 & 1 \\ & 1 & 0 & 0 \\ & 1 & 0 & 0 \end{array}$$

A_G

Fact: G has a PM iff $\mathrm{Per}(A_G)>0$

$\mathrm{Per}_n(A)=\sum_{\sigma\in S_n}\prod_{i\in[n]} A_{i\sigma(i)}$

[Jacobi'1890] PM $\in$ P (P = polynomial time)

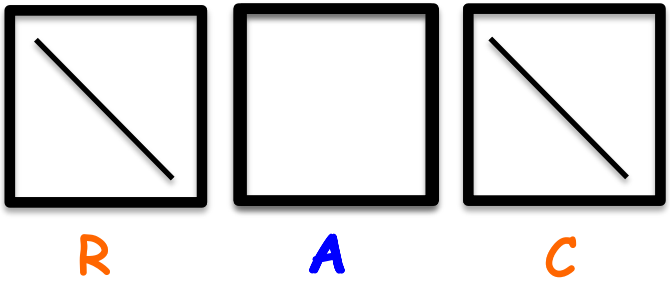
Matrix Scaling
[...Sinkhorn'64,...]

A non-negative matrix.

A doubly-stochastic (DS): $A\mathbf{1}=\mathbf{1},\ A^{t}\mathbf{1}=\mathbf{1}$

Scaling:

Multiply rows & columns by scalars



Why?
- Numerical analysis
- Signal processing
- Approx Permanent
- Perfect matching
- ......

DS-Scaling:

Find (if exists?) R,C *diagonal* s.t.

RAC has row-sums & col-sums ≈ 1

←→ Per(A)>0

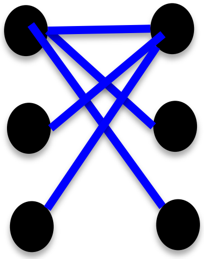
Scaling algorithm [Sinkhorn'64,…]

A non-negative matrix. Try making it doubly stochastic.
(e.g. the adjacency matrix $A=A_G$ of a bipartite graph G)

Find (if exists?) R,C *diagonal* s.t.
RAC has row-sums & col-sums ≈ 1

Hard to do simultaneously…
Let's deal with rows & cols separately!

1	1	1
1	0	0
1	0	0



Scaling algorithm [Sinkhorn'64]

Scale rows

1/3	1/3	1/3
1	0	0
1	0	0

Scaling algorithm

Scale columns

1/7	1	1
3/7	0	0
3/7	0	0

Scaling algorithm

Scale rows

1/15	7/15	7/15
1	0	0
1	0	0

Scaling algorithm

Scale columns

0	1	1
1/2	0	0
1/2	0	0

Scaling algorithm

Scale rows

0	1/2	1/2
1	0	0
1	0	0

Scaling algorithm

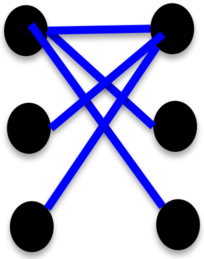
Scale columns

0	1	1
1/2	0	0
1/2	0	0

Scaling algorithm

Scale rows

0	1/2	1/2
1	0	0
1	0	0



No convergence!

No perfect matching: Per(A)=0

Scaling algorithm

A non-negative matrix. Try making it doubly stochastic.

Scaling factors

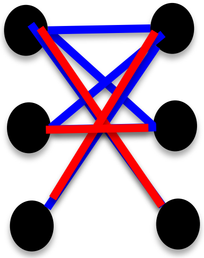
$R(A)$ = diag(row sums)$^{-1}$ $C(A)$ = diag(column sums)$^{-1}$

≠0

```
Repeat $n^3$ times:
   Scale rows $A \leftarrow R(A) \times A$
   Scale cols $A \leftarrow A \times C(A)$
```

1	1	1
1	1	0
1	0	0

"Alternating minimization" heuristic



Scaling algorithm

A non-negative matrix. Try making it doubly stochastic.

$R(A)$ = diag(row sums)$^{-1}$ $C(A)$ = diag(column sums)$^{-1}$

```
Repeat $n^3$ times:
   Scale rows $A \leftarrow R(A) \times A$
   Scale cols $A \leftarrow A \times C(A)$
```

Scale rows

1/3	1/3	1/3
1/2	1/2	0
1	0	0

Scaling algorithm

A non-negative matrix. Try making it doubly stochastic.

$R(A)$ = diag(row sums)$^{-1}$ $C(A)$ = diag(column sums)$^{-1}$

```
Repeat $n^3$ times:
   Scale rows $A \leftarrow R(A) \times A$
   Scale cols $A \leftarrow A \times C(A)$
```

Scale columns

2/11	2/5	1
3/11	3/5	0
6/11	0	0

Scaling algorithm

A non-negative matrix. Try making it doubly stochastic.

$R(A)$ = diag(row sums)$^{-1}$ $C(A)$ = diag(column sums)$^{-1}$

```
Repeat $n^3$ times:
    Scale rows $A \leftarrow R(A)\times A$
    Scale cols $A \leftarrow A\times C(A)$
```

Scale rows

10/87	22/87	55/87
15/48	33/48	0
1	0	0

Scaling algorithm

A non-negative matrix. Try making it doubly stochastic.

$R(A)$ = diag(row sums)$^{-1}$ $C(A)$ = diag(column sums)$^{-1}$

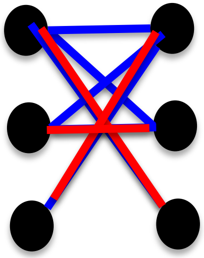
```
Repeat $n^3$ times:
   Scale rows $A \leftarrow R(A) \times A$
   Scale cols $A \leftarrow A \times C(A)$
```

0	0	1
0	1	0
1	0	0

Scale rows

Converges!

Has perfect matching: Per(A)>0



Analysis of the algorithm

[Linial-Samorodnitsky-W'01]

A non-negative (0,1) matrix.

```
Repeat $t=n^3$ times:
   Scale rows $A \leftarrow R(A) \times A$
   Scale cols $A \leftarrow A \times C(A)$

Test if $A_t \approx DS$ (up to $1/n$)
   Yes: Per($A$) > 0.
   No:  Per($A$) = 0.
```

0	0	1
		0
1	0	0

Algorithm for Perfect Matching

Analysis: Per(A_i) a progress measure!

- Per(A_i) ≤ 1 (easy)
- Per(A_i) grows* by $(1+1/n)$ (AMGM)
- Per(A)>0 ➔ Per(A_1)> $1/n^n$ (easy)

Non-uniform scaling

Given: A non-negative matrix, p,q vectors.

Task: Scale it so that it has these row and columns sums, namely so that $A1 \approx p$, $1A \approx q$ (if possible)

- Related to max-flow in graphs
- Further generalized to the marginal problem: Scale a multivariate distribution to have some given marginals

The same Alternating Minimization algorithm works!

Done (baby case):

Bip matching & Matrix Scaling

Now (real thing):

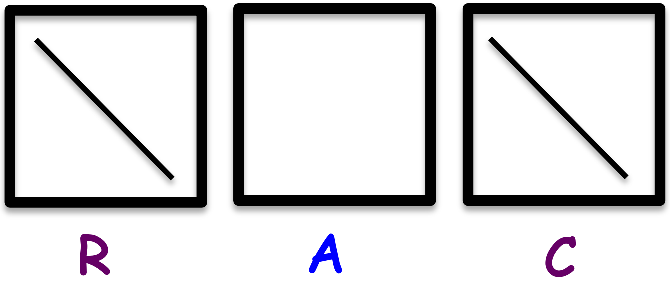
NC-SING & Operator Scaling

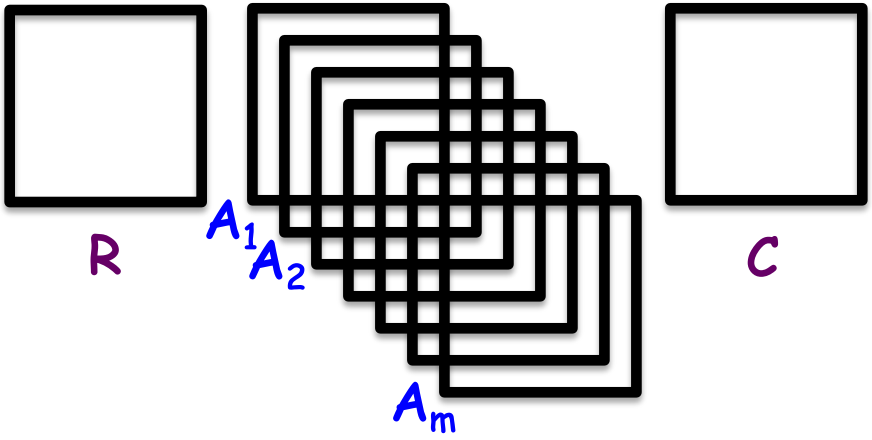
[Gurvits'04]

[Garg,Gurvits,Oliveira,W'15]

[Gurvits '04] Quantum leap

	Matrix Scaling	Operator Scaling
Input	Positive matrix	Positive operator
Norm	L_1	L_2
R,C	Diagonal	Invertible
DS	$A1=1$, $A^\dagger 1=1$	$\sum_i A_i A_i^\dagger = I$ $\sum_i A_i^\dagger A_i = I$
	$A1=p$, $A^\dagger 1=q$	$\sum_i A_i A_i^\dagger = P$ $\sum_i A_i^\dagger A_i = Q$





Operator Scaling [Gurvits '04]

a quantum leap

Algebra

Input: $L=(A_1,A_2,\dots,A_m)$

Symbolic matrix

L: $A_1x_1+A_2x_2+\dots+A_mx_m$

Is L C-singular? ←

Is L NC-singular? ↔ [GGOW'15]

Quantum Inf. Theory

Input: $L=(A_1,A_2,\dots,A_m)$

Completely positive operator

$L(P)=\sum_i A_i P A_i^\dagger$

P psd → $L(P)$ psd

L doubly stochastic:

$\sum_i A_i A_i^\dagger = I$ $\quad$ $\sum_i A_i^\dagger A_i = I$

$L(I)=I$ $\quad$ $L^\dagger(I)=I$

Can we (not) scale L?

Operator scaling algorithm

[Gurvits '04, Garg-Gurvits-Olivera-W'15]

$L=(A_1,A_2,\dots,A_m)$.

Scaling: $L \rightarrow RLC$, R,C invertible, DS: $\sum_i A_i A_i^\dagger = I \quad \sum_i A_i^\dagger A_i = I$

Scaling factors: $R(L) = (\sum_i A_i A_i^\dagger)^{-1/2} \quad C(L) = (\sum_i A_i^\dagger A_i)^{-1/2}$

```
Repeat $\boxed{t=n^c}$ times:
   Scale "rows" $L \leftarrow R(L) \times L$
   Scale "cols" $L \leftarrow L \times C(L)$
Test if $L_t \approx DS$ (up to $1/n$)
   Yes: $L$ NC-nonsing      $cap(L)>0$
   No:  $L$ NC-singular     $cap(L)=0$
```

Progress measure

Capcity(L) = $\inf_{P>0} \det(L(P))/\det(P)$

Algorithm: Group action
Measure: "Invariant"
Analysis: Degrees of invariant polynomials

Analysis:
- $Cap(L_i) \leq 1$ asy)
- $Cap(L_i)$ grows* by $(1+1/n)$ (AMGM)
- $Cap(L)>0 \rightarrow Cap(L_1)>\exp(-n^c)$ [GGOW'15]

6 areas, 6 problems [GGOW'15+16]

$L=(A_1,A_2,\dots,A_m)$, $A_i \in \mathrm{Mat}_n(F)$ (e.g. $F=Q$)

Linear algebra $A_i: F^n \to F^n$ linear maps

Q1: $\exists$ subspace U s.t. $\dim(\mathrm{span}\{A_iU\}_i) < \dim(U)$? In P

Arithmetic complexity theory: A describes a polynomial:

Q2: $L(X) = \det(\sum_i A_i \times X_i) = 0$? In P

Quantum Information Theory L positive operator: $L(P) = \sum_i A_i P A_i^\dagger$

Q3: $\inf_{P\succ 0} \det(L(P))/\det(P) = 0$? In P

Non-commutative Algebra

Q4: Is $L(x)=\sum_i A_i x_i$ singular in $F\langle(x)\rangle$? [word problem] In P

Invariant Theory L orbit of $G = SL_n(F)\times SL_n(F)$

Q5: Is $0 \in \underline{O_G(L)}$? [null cone problem] In P

Analysis $A_i: R^n \to R^n$ linear maps

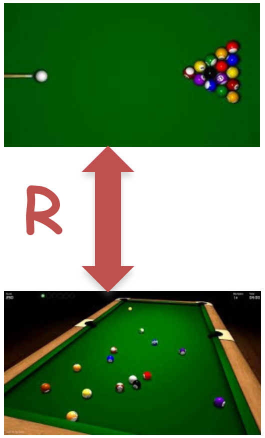
Q6: $\exists\, C<\infty\ \forall f_i: R^n \to R_+ \ \int_{x\in R^n} (\Pi_i f_i(A_i x)) \le C\, \Pi_j |f_i|_m$? In P

Q1-Q5 are equivalent! Q6 special case

Here: Linear groups* (of matrices) act* on vector spaces (over $\mathbb{C}$)

Algebraic: Polynomial invariants

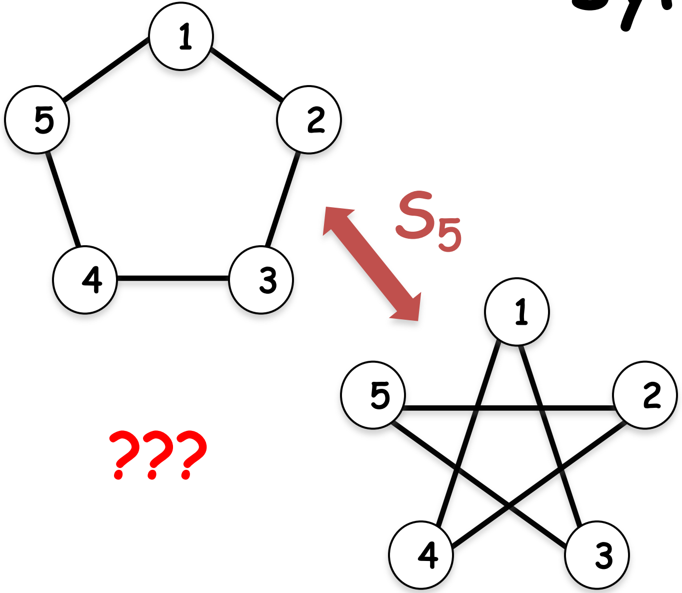
Geometric: Non-commutative duality



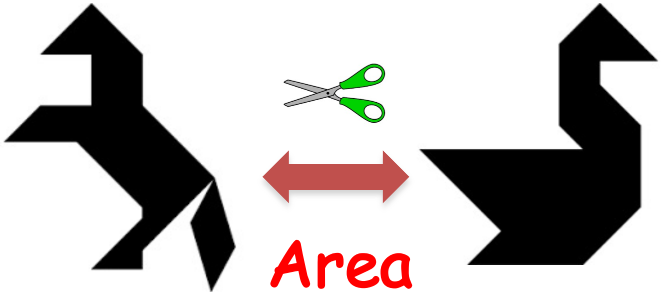
- Energy
- Momentum

Invariant Theory

symmetries, group actions, orbits, invariants



???



Invariant theory

Nullcone Membership:

Given v, does $v \in N(G)$?

Dual to Scaling problems!

Captures numerous problems across Math, CS, Physics, for different group actions

G acts on $V=F^k$, and se =ℂ)

Orbit: Gv = {

Invariant

V^G = { p ∈

Ex1:

V^G = (G)= {0}

Ex2: G

V^G = < det(rices}

[Hilbert] Invaria ely generated!

Degree bounds?

Key to alg analysis

Algebraic Variety

Nullcone: N(G)= {v : p(v)=0 for all $p \in V^G$, deg(p)>0}

[Hilbert,Mumford] $v \in N(G)$ ←→ $0 \in \overline{Gv}$ ←→ $\inf_{g \in G} |gv|=0$

Analytic ←→ Algebraic

Unification and generalization I

[BGOWW'17,F'17,BFGOWW'18]

Alternate minimization on groups

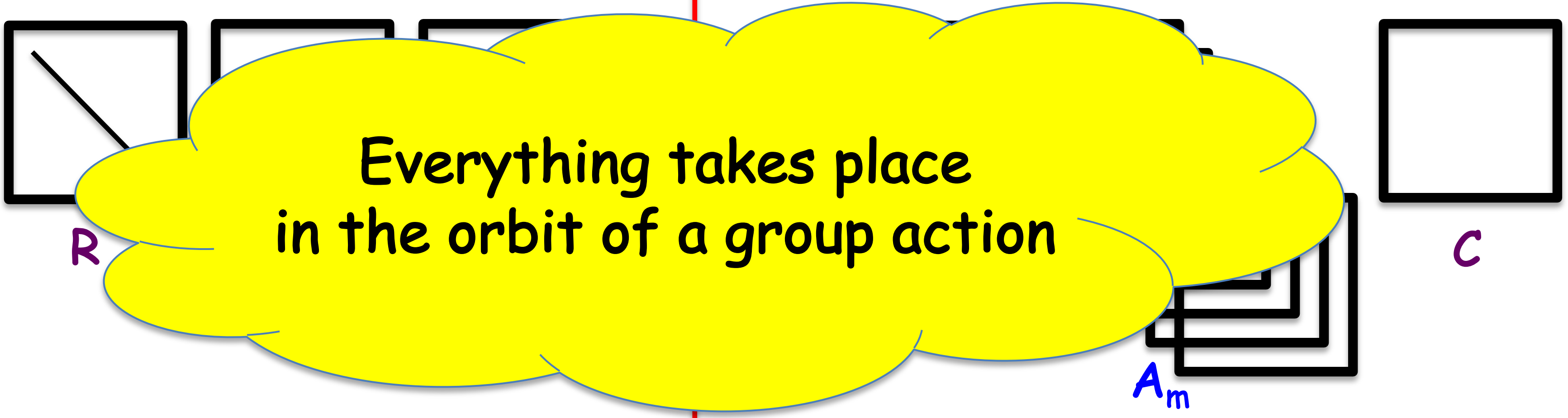
[BGOWW'17,F'17,BFGOWW'18]

Goal: Matrix Scaling

$A\mathbf{1}=\mathbf{1}$, $A^{t}\mathbf{1}=\mathbf{1}$

Operator Scaling

$\sum_i A_i A_i^{t} = I$ $\sum_i A_i^{t} A_i = I$



Everything takes place
in the orbit of a group action

Alg: Alt. min. $T_n \times T_n$
(Diagonal group)^2
action on matrices

Alg: Alt. min. $GL_n \times GL_n$
(General linear group)^2
action on tensors

Analysis: minimizing a potential function (permanent, capacity)

Alternate Minimization

numerous other examples

(statistics, optimization, sampling, machine learning,…)

“solve” $f(z_1,z_2,\dots,z_i,\dots,z_k)$ all z_i complex

“solve” $f(a_1,a_2,\dots,z_i,\dots,a_k)$ one z_i simple/local

(a_j fixed)

Some examples we don’t understand well…

Alternate minimization on groups

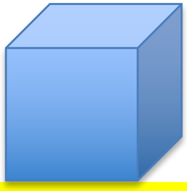
Alt Minimization (coordinate descent)

(statistics, optimization, machine learning,….)

"solve" $f(z_1,z_2,\ldots,z_i,\ldots,z_k)$ all z_i complex

"solve" $f(a_1,a_2,\ldots,z_i,\ldots,a_k)$ one z_i simple/local

Here: group-theoretic framework



k-tensor

$G = G_1 \times G_2 \times \ldots\ldots. \times G_k$ $G_i = SL_n(C)$ or $ST_n(C)$

$V = V_1 \otimes V_2 \otimes \ldots\ldots \otimes V_k$ $V_i = C^n$, G_i acts on i-fibers of V

Non-convex

Goal: Given $v \in V$, scale it (make all "marginals" uniform)

[THM] Alt Min: $|v' - \text{"scaled"}| < \varepsilon$ in poly($|v|$, n, $1/\varepsilon$) steps.

Alternate Minimization over groups

Applications and analysis

$G = G_1 \times G_2 \times \ldots\ldots \times G_k$ G_i = $SL_n(C)$ or $ST_n(C)$

k-tensor

$V = V_1 \otimes V_2 \otimes \ldots \otimes V$ … of V

Goal … (G)

Why does such a simple greedy algorithm converge?

What connects scaling and nullcone problems?

Non-product groups? (no alternate minimization)

es

h!

y)

ε) steps.

y)

(1+ ε/n)   (AMGM)

- v∉N(G) → |v|>exp(-poly(n/ε)) (inv+rep th.)

Same alg…

“same” analysis.

Potential: $|.|_2$

Alternate Minimization over groups
Analysis from invariants polynomials

Must prove $v \notin N(G)$ → $|v| > \exp(-n^c)$ [a "diameter" bound]
Invariant Theory: old tools + new bounds

[Hilbert,Mumford] $v \in N(G)$ ↔
$p(v)=0$ $\forall$ invariant polynomials p [$p(v)=p(gv)$ $\forall g \in G$]

Doubly exp algs

$v \notin N(G)$ → $\exists$ invariant integer polynomial p s.t. $p(v) \neq 0$
degree$(p)=d$, height$(p)=h$ → $1 \leq |p(v)| \leq d^{O(n)} h |v|^d$

Analysis only

[Derksen] V^G is generated in degree $d < \exp(n^2)$

[Cayley] Omega Process: generating invariants of SLn actions of any degree d → height $h < d^{O(d)}$

Next talk – some highlights

- Non-commutative duality (extending LP duality)
- Moment map (extending Euclidean gradients)
- Geodesic convexity (extending Euclidean one)
- Non-commutative 1st & 2nd order (geodesic) algs
- Analysis via Invariant Theory and Representation Th.
- Moment polytopes
- …

Conclusions & Open Problems

General themes

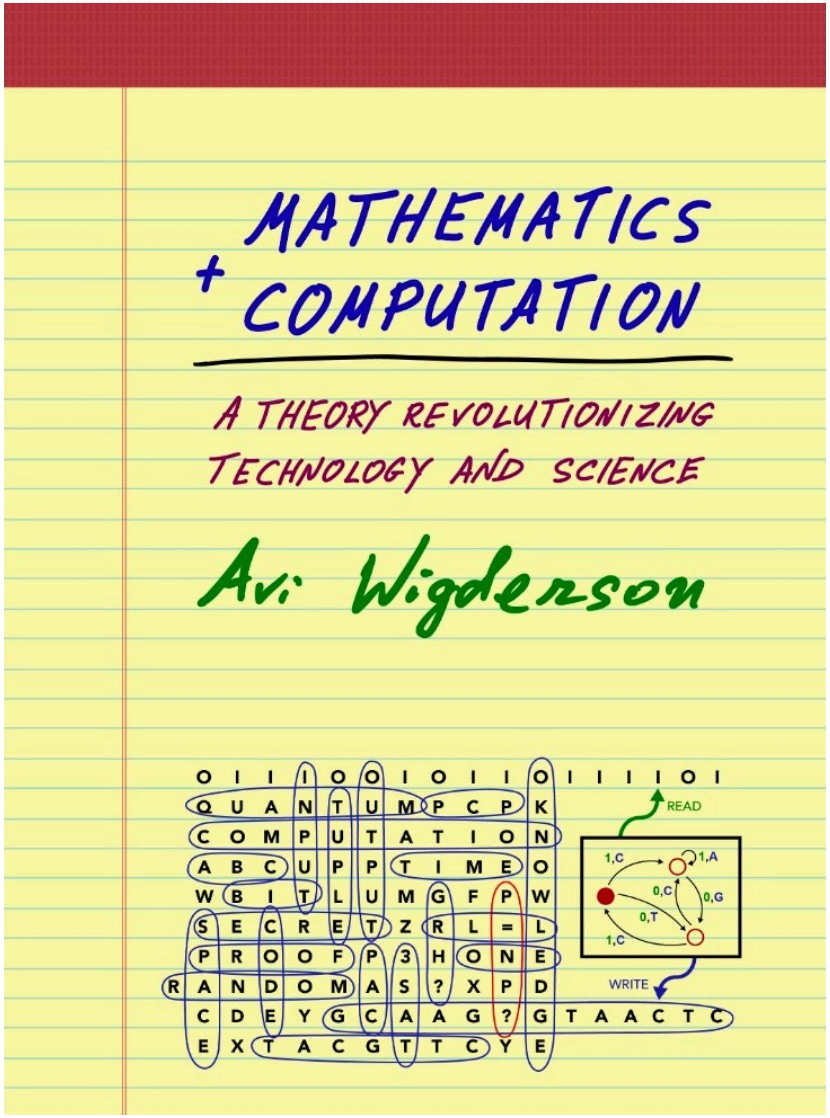
- Algorithms & complexity interacts with Math
- Analytic solutions to algebraic problems
- Algebraic analysis of continuous algorithms
- Symmetry is prevalent, using it is powerful

Natural research directions

- Algs still exponential for some applications
- Power of geodesic algs for comb. optimization
- Nullcone problems abound. Nullcone $\in$ P?
- C-SING $\in$ P? "P vs. NP"? Any lower bounds??

  [Makam-W'19] C-SING is *not* a nullcone problem!

Book ad



- Published by Princeton University Press

- Free (forever) on my website

- Comments welcome!