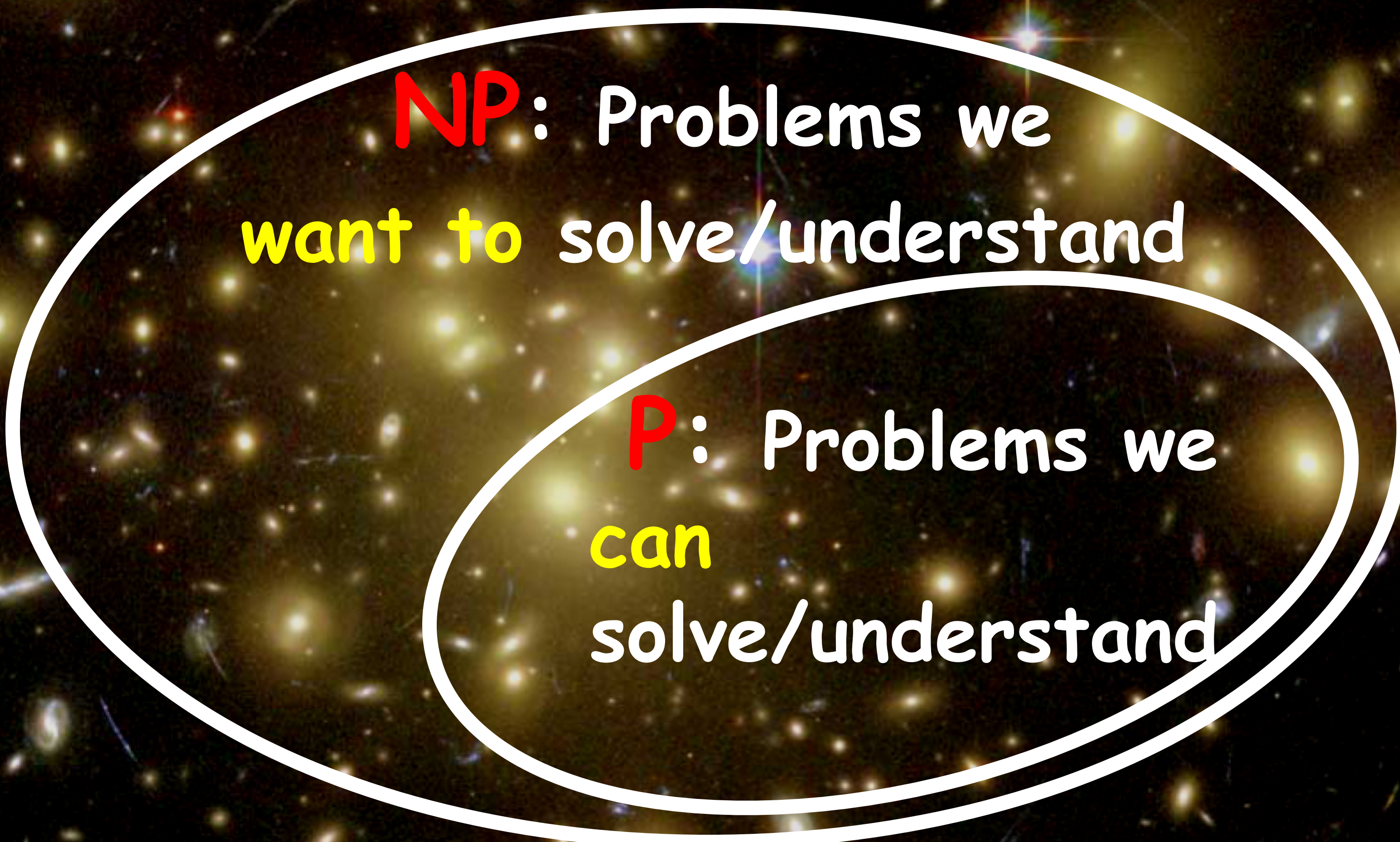
Optimization, Complexity, and Math

(or, can we prove $P \neq NP$ using gradient descent?)

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

Scientific / Mathematical/ Intellectual / Computational problems



P=NP?

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
Nullcone membership & orbit closure intersection

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
Symbolic matrices, algebraic identities, lower bounds

Optimization
Efficiently solving certain general families of
- Quadratic systems of equations
- Exponentially large linear LPs

Optimization, Complexity and Math
through one problem and one algorithm

One problem
Singularity of Symbolic Matrices

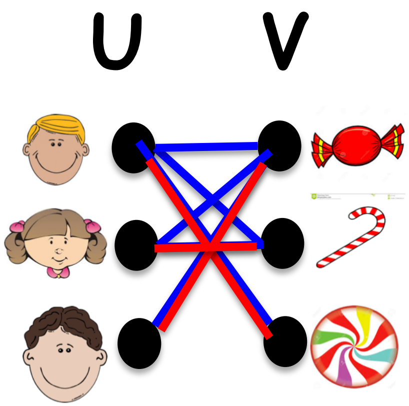
One algorithm
Alternating minimization

The problem(s)

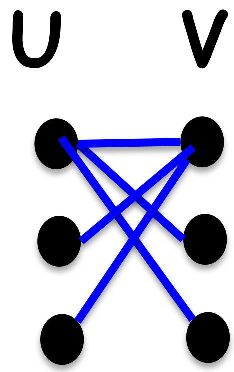
Perfect Matchings (PMs)

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

$|U|=|V|=n$



G'



G

V

U			
	1	1	1
	1	0	0
	1	0	0

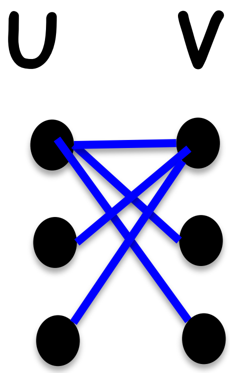
A_G

Fact: G has a PM iff $Per(A_G)>0$

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

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

PMs & symbolic matrices [Edmonds'67]



G

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 $\mathrm{Det}(A_G(X)) \neq 0$ ($\in$ P)

Symbolic matrices [Edmonds'67]

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

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

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

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

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

$L(X)$

[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

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

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

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

Input: $A_1, A_2, \dots, A_m \in M_n(F)$

SING : Is $L(X)$ singular?

x_i commute

in $F(x_1, x_2, \dots, x_m)$

[Lovasz '79] SING $\in$ RP

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

x_i do not commute

in $F<(x_1, x_2, \dots, x_m)>$ (free skew field)

[Cohn'75] NC-SING Decidable

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

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

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

The algorithm

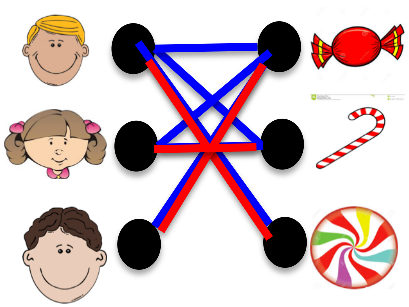
Alternate minimization

Perfect Matchings (PMs)

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

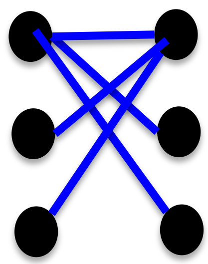
$|U|=|V|=n$

U V



G'

U V



G

V (columns), U (rows):

1	1	1
1	0	0
1	0	0

A_G

Fact: G has a PM iff $Per(A_G)>0$

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

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

Matrix Scaling

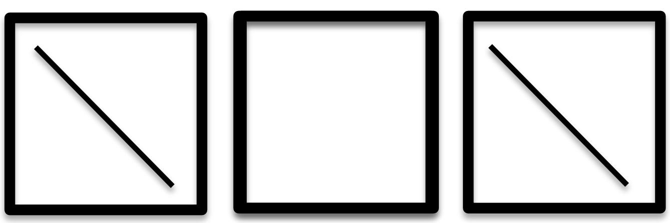
[...Sinkhorn'64,...]

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



R A C

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

$\Leftrightarrow$ Per(A)>0

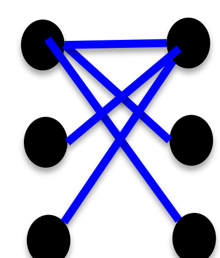
Scaling algorithm [Sinkhorn'64,…]

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

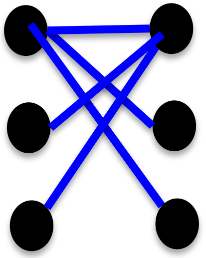
Scale columns

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.

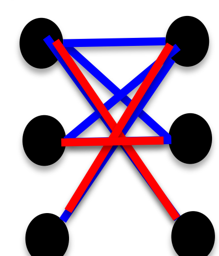
Scaling factors

$R(A)$ = 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 n³ times:
  Scale rows A ← R(A)×A
  Scale cols A ← A×C(A)
```

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

Scale rows

Scaling algorithm

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

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

```
Repeat n³ times:
   Scale rows A ← R(A)×A
   Scale cols A ← A×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 n³ times:
  Scale rows A ← R(A)×A
  Scale cols A ← A×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)$^{-1}$

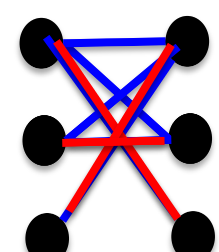
```
Repeat n³ times:
   Scale rows A ← R(A)×A
   Scale cols A ← A×C(A)
```

0	0	1
0	1	0
1	0	0

Scale rows

Converges!

Has perfect matching: Per(A)>0



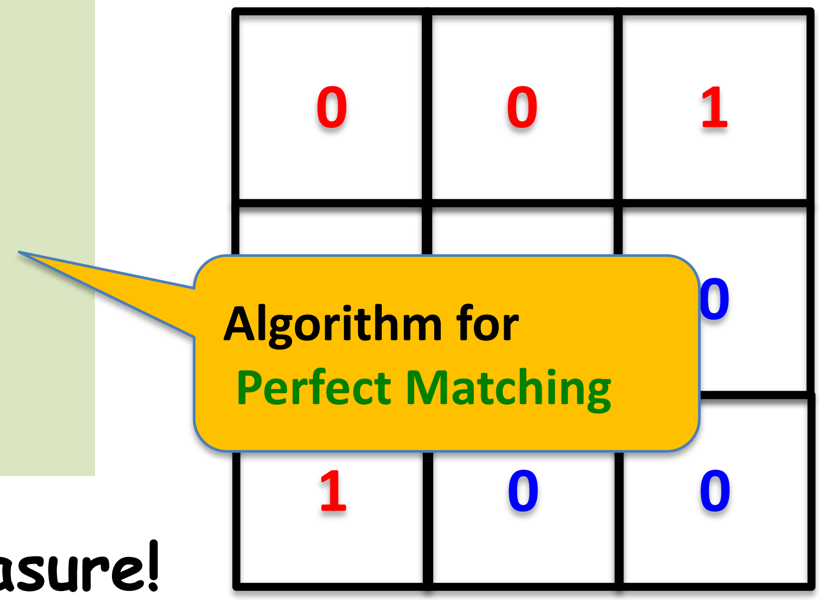
Analysis of the algorithm

[Linial-Samorodnitsky-W'01]

A non-negative (0,1) matrix.

```
Repeat t=n³ times:
   Scale rows A ← R(A)×A
   Scale cols A ← A× C(A)

Test if Aₜ≈DS (up to 1/n)
   Yes: Per(A) > 0.
   No:  Per(A) = 0.
```



Analysis: $Per(A_i)$ a progress measure!

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

Done (baby case):

Bip matching & Matrix Scaling

Now (real thing):

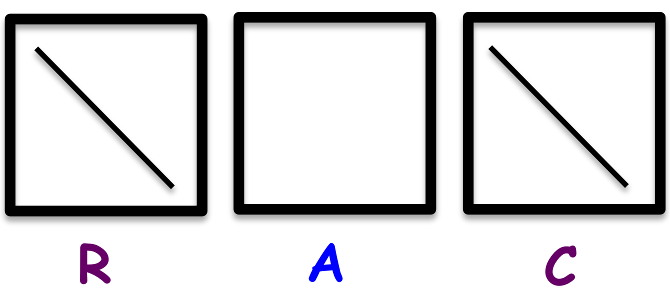
NC-SING & Operator Scaling

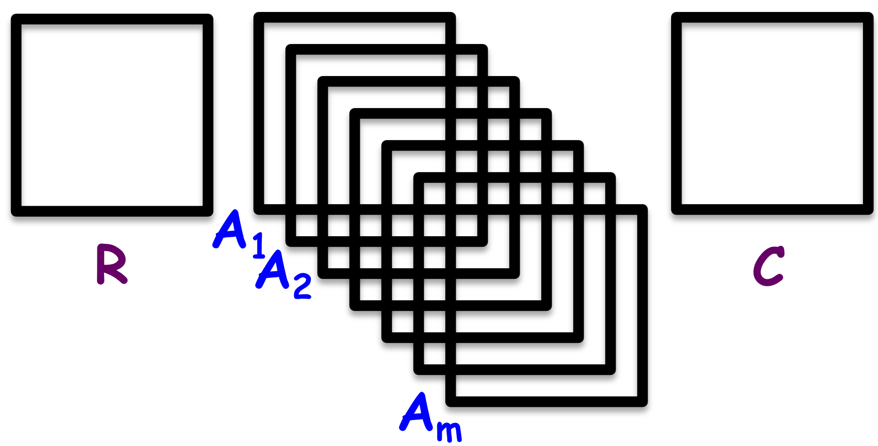
[Gurvits'04]

[Garg,Gurvits,Oliveira,W'15]

[Gurvits '04] Quantum leap

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





Operator Scaling [Gurvits '04]

a quantum leap

Algebra

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

Symbolic matrix

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

Quantum Inf. Theory

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

Completely positive operator

$L(P)=\sum_i A_iPA_i^{\dagger}$

P psd $\rightarrow L(P)$ psd

L doubly stochastic:

$\sum_i A_iA_i^{\dagger}=I$ $\quad$ $\sum_i A_i^{\dagger}A_i=I$

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

Is L C-singular? $\leftarrow$

Is L NC-singular? $\leftrightarrow$ [GGOW'15]

Can we (not) scale L?

Operator scaling algorithm

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

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

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

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

```
Repeat t=n^c times:
   Scale "rows" L ← R(L)×L
   Scale "cols" L ← L× C(L)
Test if L_t≈DS (up to 1/n)
   Yes:L NC-nonsing
   No: L NC-singular
```

$cap(L)>0$

$cap(L)=0$

Progress measure

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

Algorithm: Group action

Measure: "Invariant"

Analysis: Degrees of invariant polynomials

Analysis:
- $Cap(L_i) \leq 1$ sy)
- $Cap(L_i)$ 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]

$L=(A_1,A_2,\dots,A_m)$, $A_i \in Mat_n(F)$ (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: $\inf_{P>0} \det(L(P))/\det(P) = 0$? **In P**

Non-commutative Algebra

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

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

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

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

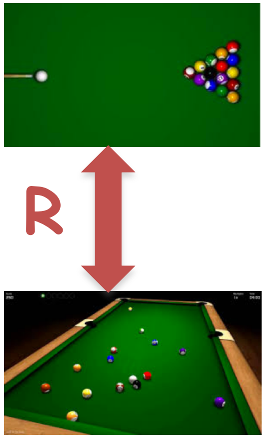
Q6: $\exists\, C<\infty\ \forall f_i: R^n \to R_+ \quad \int_{x\in R^n} (\prod_i f_i(A_i x)) \le C \prod_j |f_i|_m$? **In P**

Q1-Q5 are equivalent! Q6 special case

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

Algebraic: Polynomial invariants

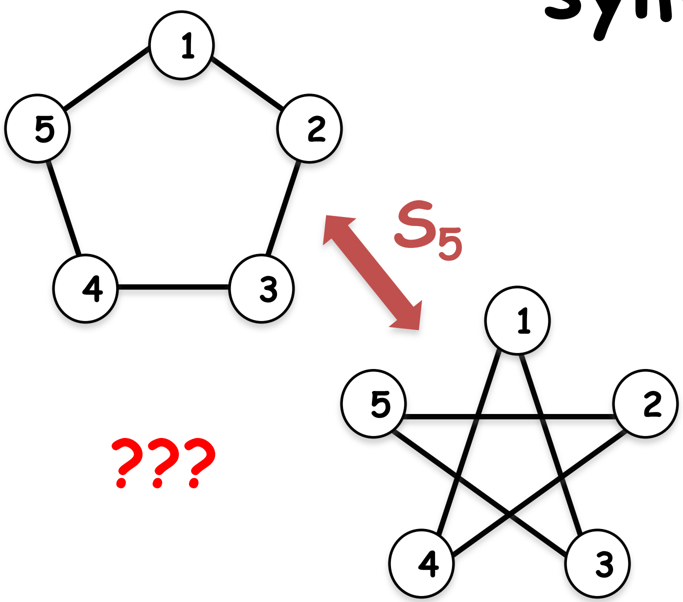
Geometric: Non-commutative duality

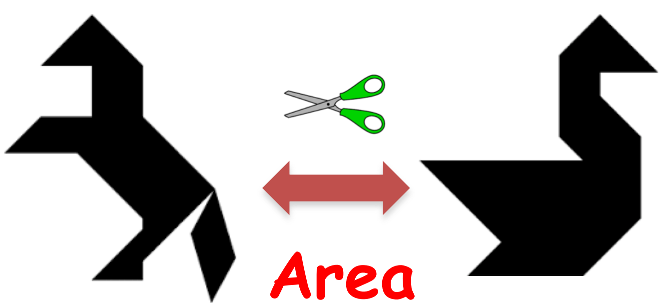


- Energy
- Momentum

Invariant Theory

symmetries, group actions, orbits, invariants





Invariant theory

G acts on $V=F^k$, and ... ($F=\mathbb{C}$)
Orbit: Gv = ... Gv
Invariant ...
V^G = { p ...

Nullcone Membership:

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

Dual to Scaling problems!

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

Ex1: ... tes
V^G ... (G)= {0}

Ex2: ... RvC
V^G = < ... atrices}

[Hilbert] Invari... 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

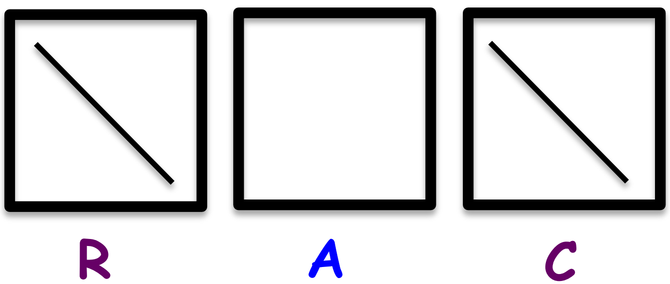
[BGOWW'17,F'17,BFGOWW'18]

Alternate minimization on groups

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

Goal: Matrix Scaling

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



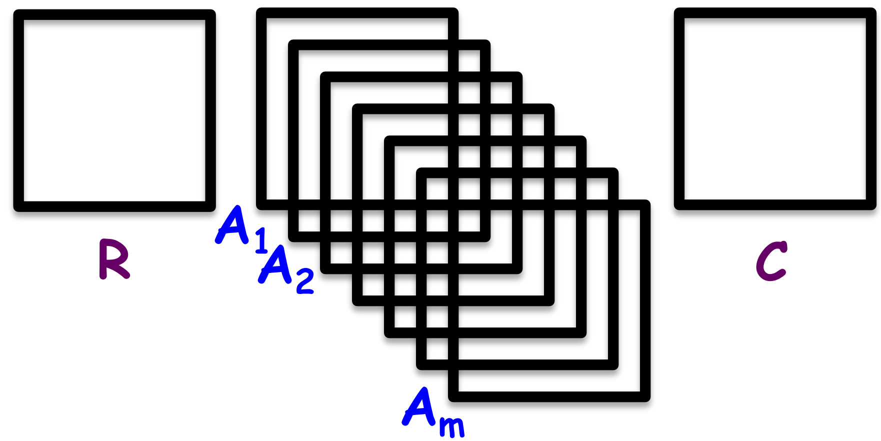
Alg: Alt. min. $T_n \times T_n$

(Diagonal group)^2

action on matrices

Operator Scaling

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



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, 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

(a_j fixed)

Some examples we don't understand well…

Alternate minimization on groups

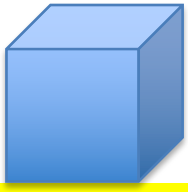
Alt Minimization (coordinate descent)

(statistics, optimization, machine learning,…)

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

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

Here: group-theoretic framework



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

k-tensor

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

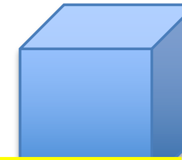
Non-convex

Goal: Given $v \in V$, scale it

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

Alternate Minimization over groups

Applications ...



$G = G_1 \times$ ...

$V = V$ ... f V

G ... N(G)

-

-

- ... ory)

- Any ... s)

Why does such a simple greedy algorithm converge?

What connects scaling and nullcone problems?

Non-product groups? (no alternate minimization)

... s ... ough!

[THM] Alt Min ... n poly(|v|,n,1/ε) steps.

Same alg, "same" analysis.

Potential: $|.|_2$

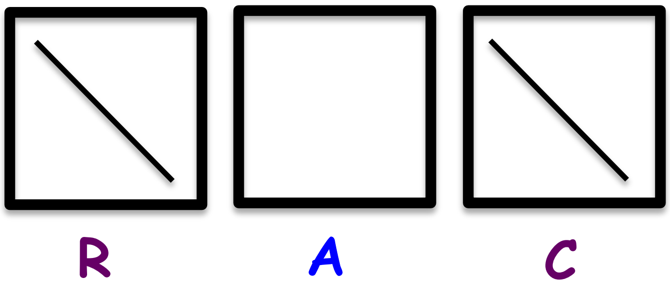
- $|v_i| \leq 1$ (easy)
- $|v_i|$ grows* by $(1+1/n)$ (AMGM)
- $v \notin N(G) \rightarrow |v| > \exp(-n^c)$ (inv+rep th. ++)

Unification & generalization II

Non-commutative duality,
Geodesic convexity,
Moment map

Goal: Matrix Scaling

$A\mathbf{1}=\mathbf{1}, \; A^{\dagger}\mathbf{1}=\mathbf{1}$



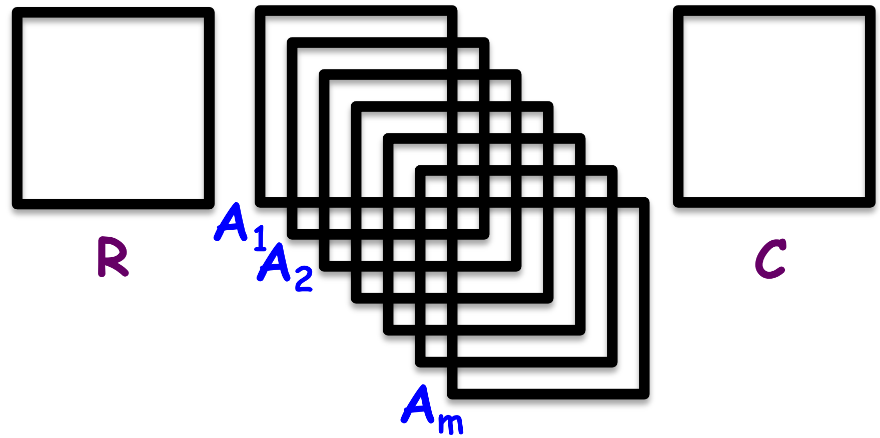
Alg: Alt. min. $T_n \times T_n$

(Diagonal group)^2

action on matrices

Operator Scaling

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



Alg: Alt. min. $GL_n \times GL_n$

(General linear group)^2

action on tensors

Analysis: minimize a potential function (permanent, capacity, L_2)

Let's switch analysis with goal!

Non-commutative duality

G acts on V   (G matrix group, V vector space)

Optimization

Non-convex

Goal: given v∈V compute f(v) = $\inf_{g \in G} |gv|_2$ =0?

Nullcone problem

(find the li

How to ... 0!

[Kemp ... v*≠0

μ m ... nt

- M ... l to

mini

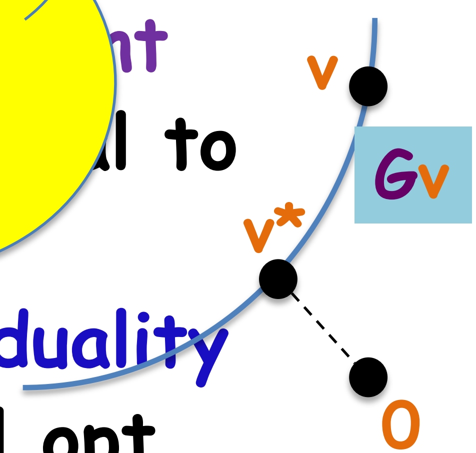
- No ... P duality

Geodesic c ... pt = global opt

Extends Euclidean convexity to Riemannian manifolds

Intuitively explains the convergence of a local, greedy algorithm like Alternate Minimization

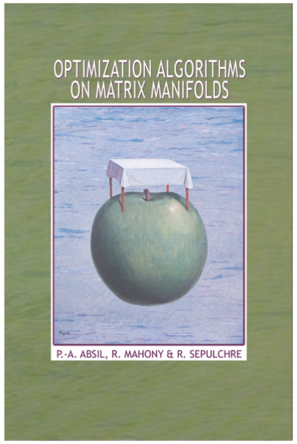
Suggests extending convex optimization tools to geodesic setting

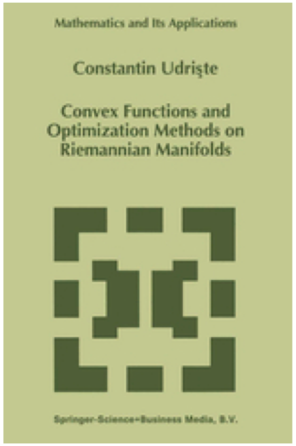


Beyond Euclidean convexity

Geodesic convexity

and algorithms





New: Quantitative complexity analysis

Geodesic optimization [AGLOW'17, BFGOWW'19]

General groups, beyond alt min & convex opt

G acts on V. Given $v \in V$, $\varepsilon > 0$, approximate the minimum of $|u|$ or $|\mu_G(u)|$ in Gv.

Different problems

Oracle access to action + function + derivatives.

First order method ("gradient descent"):
ε-convergence in $O(\mathrm{poly}(1/\varepsilon))$ iterations

Second order method ("box-constrained" Newton):
ε-convergence in $O(\mathrm{polylog}(1/\varepsilon))$ iterations

Simple parameters

$O(.)$ depends on "smoothness" of the action
(geometry of the weights of irreps of the action)

Conclusions & Open Problems

General themes

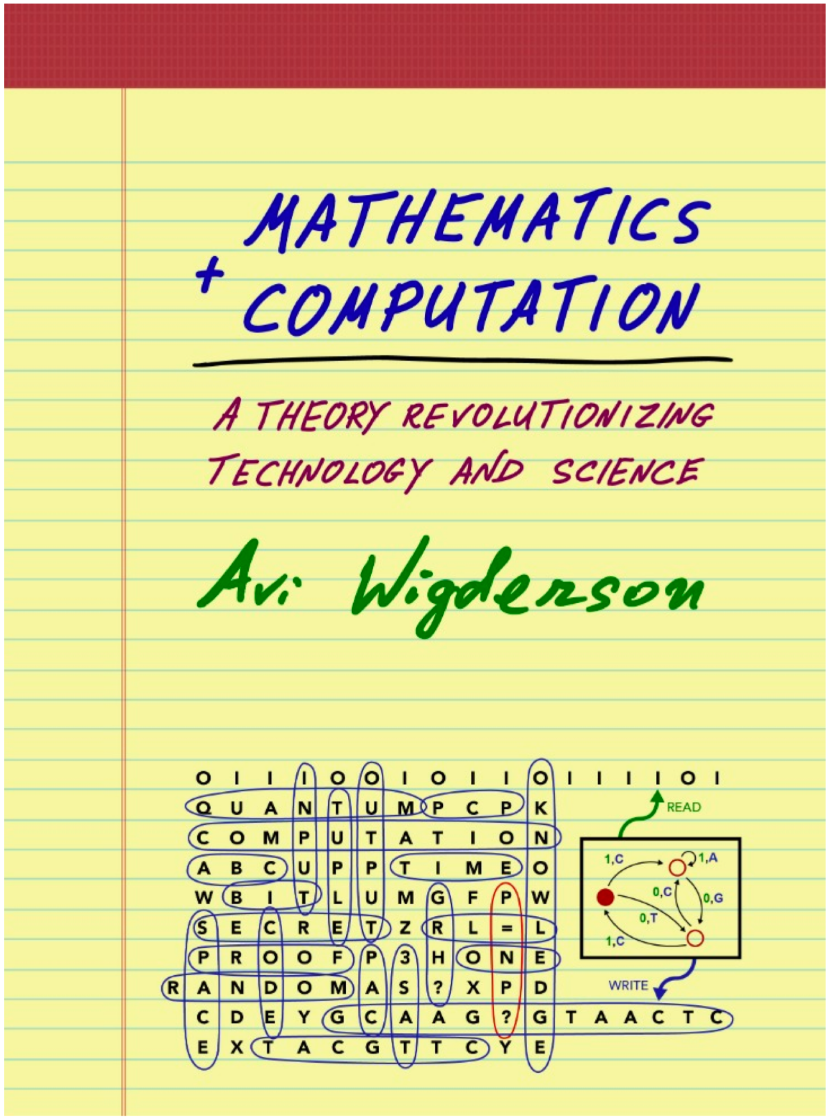
- Algorithms & complexity interacts with Math
- Analytic solutions to algebraic problems
- Algebraic analysis of continuous algorithms
- Symmetry is prevalent, using it is powerful

Natural research directions

- Algs still exponential for some applications
- Power of geodesic algs for comb. optimization
- Nulcone 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!