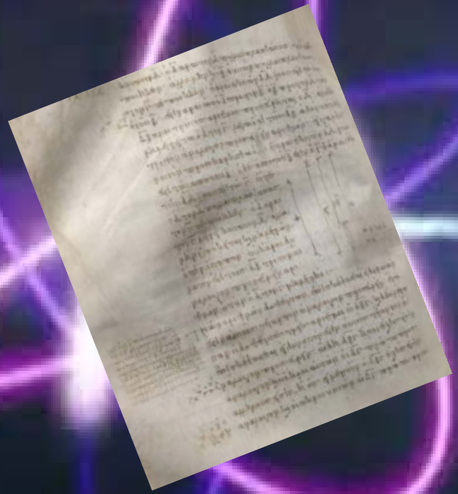
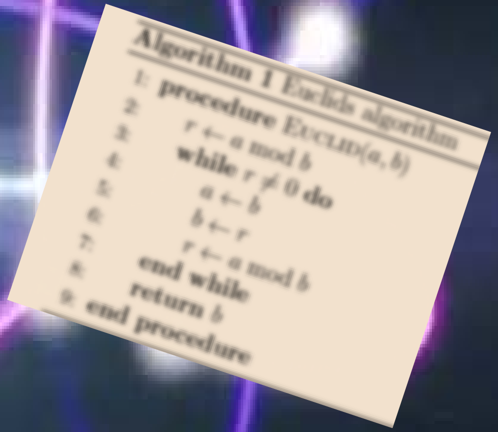
Entangling Algorithms and Proofs

|Boaz⟩ ⟨Barak|





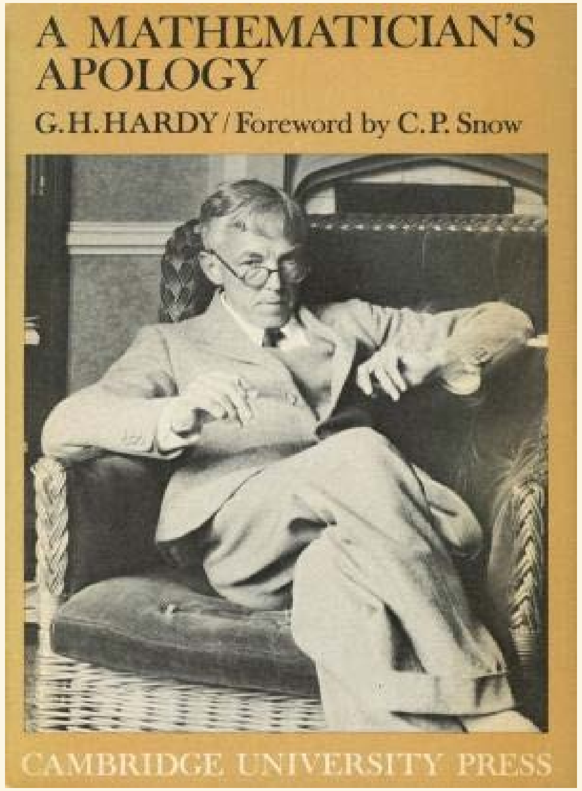
Algorithm 1 Euclid's algorithm

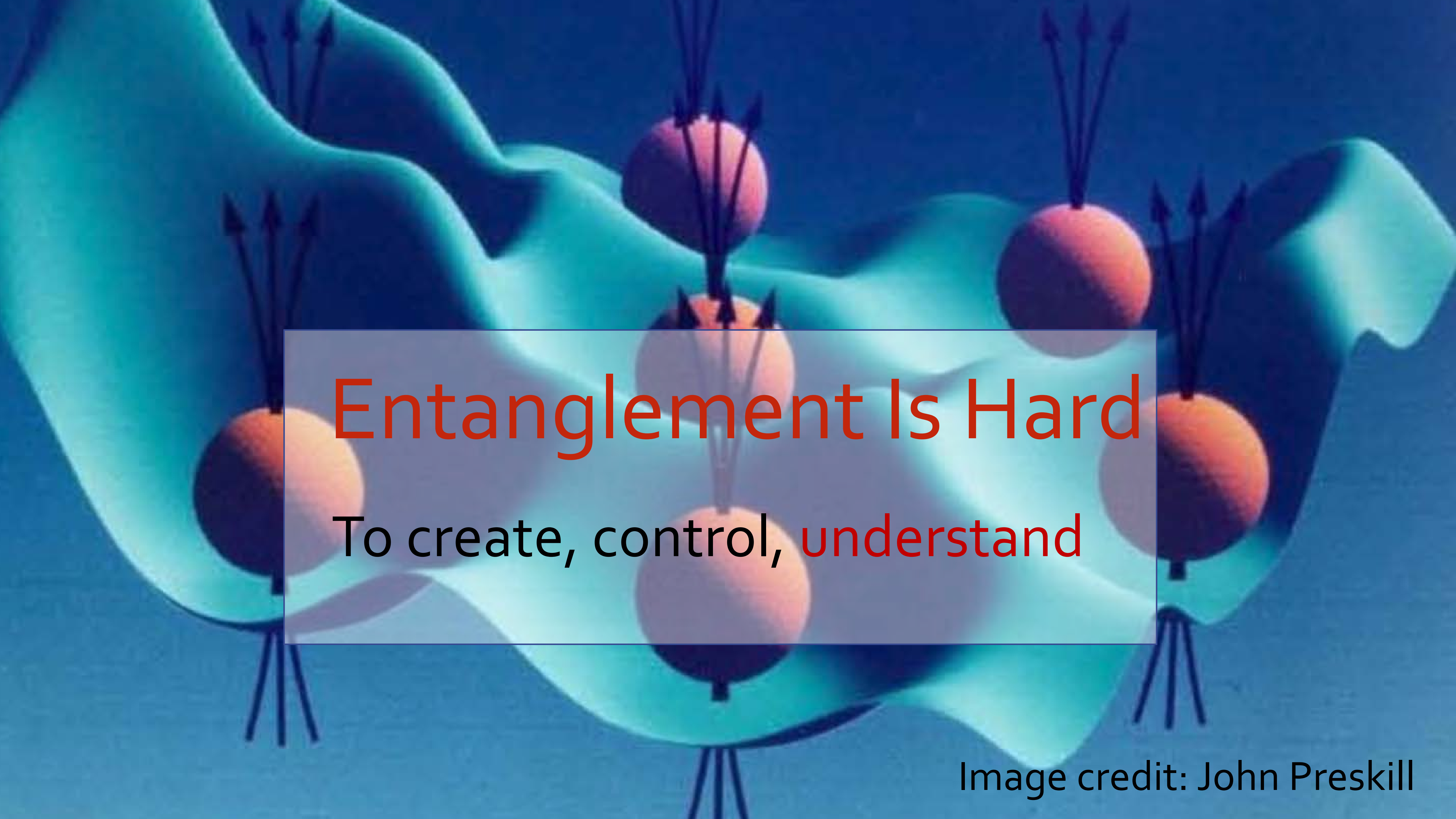
```
procedure Euclid(a, b)
    r ← a mod b
    while r ≠ 0 do
        a ← b
        b ← r
        r ← a mod b
    end while
    return b
end procedure
```

Based on joint work with Pravesh Kothari and David Steurer (arxiv 1701.06321)

QIP 2018

A classical computer scientist's apology





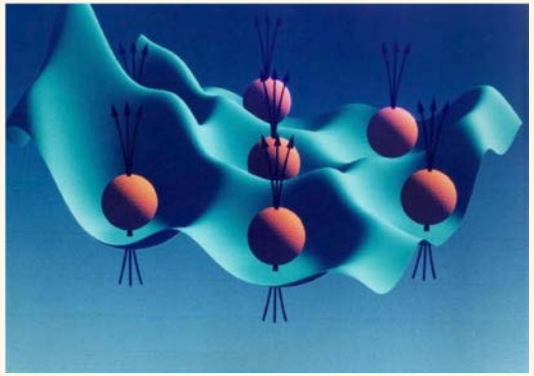
Entanglement Is Hard

To create, control, understand

Image credit: John Preskill

Entanglement Is Hard

To create, control, understand



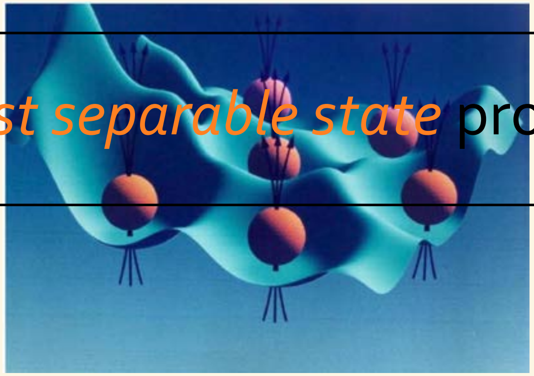
No “simple” formula for entanglement of two qudit mixed state ρ over $\mathbb{C}^{d^2}$

No “simple” formula for entanglement of e-states of $d^2 \times d^2$ measurement M.

Best known algorithms require “brute force” (i.e., $2^{\Omega(d)}$ time)

This talk: Better than brute force algorithm for (one version of) second problem.

Entanglement is Hard



Thm: *Better than brute force* (i.e. $2^{\tilde{O}(\sqrt{d})}$ time) algorithm for *best separable state* problem.

To create, control, understand

No "simple" formula for entanglement of two qudit mixed state ρ over $\mathbb{C}^{d^2}$

No "simple" formula for entanglement of e-states of $d^2 \times d^2$ measurement M.

Best known algorithms require "brute force" (i.e., $2^{\Omega(d)}$ time)

This talk: Better than brute force algorithm for (one version of) second problem.

Thm: *Better than brute force* (i.e. $2^{\tilde{O}(\sqrt{d})}$ time) algorithm for *best separable state* problem.

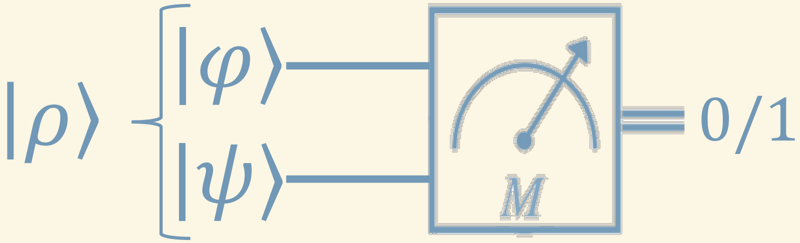
Input: Measurement M on a two qudit system ($0 \preccurlyeq M \preccurlyeq I$ is $d^2 \times d^2$ matrix)

Goal: Distinguish between:

(i) M accepts *some* separable state with probability 1

vs.

(ii) *Every* separable state is accepted with probability $\leq 1 - \epsilon$

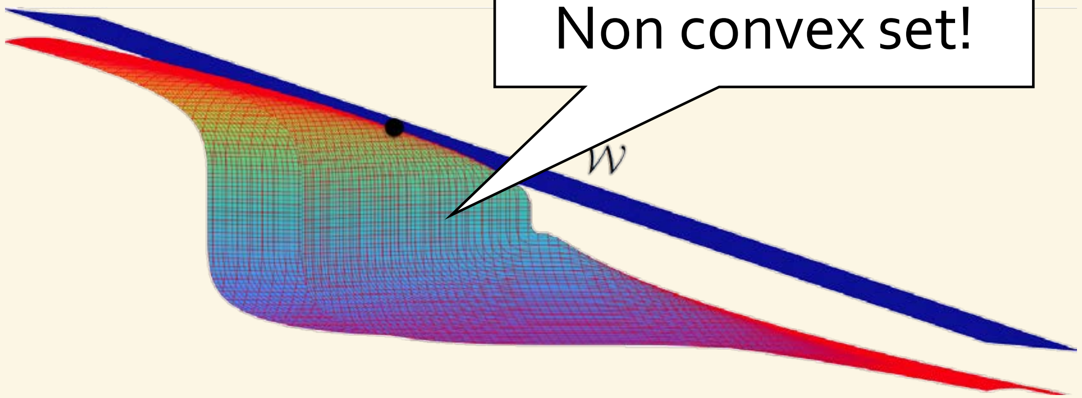


Certify M is entanglement witness

Separable states: Generated by rank one pure states $|u\rangle\langle v| \in \mathbb{C}^{d^2}$

Top e-space of M: Linear subspace $W \subseteq \mathbb{C}^{d^2}$

Goal: Find out if they intersect



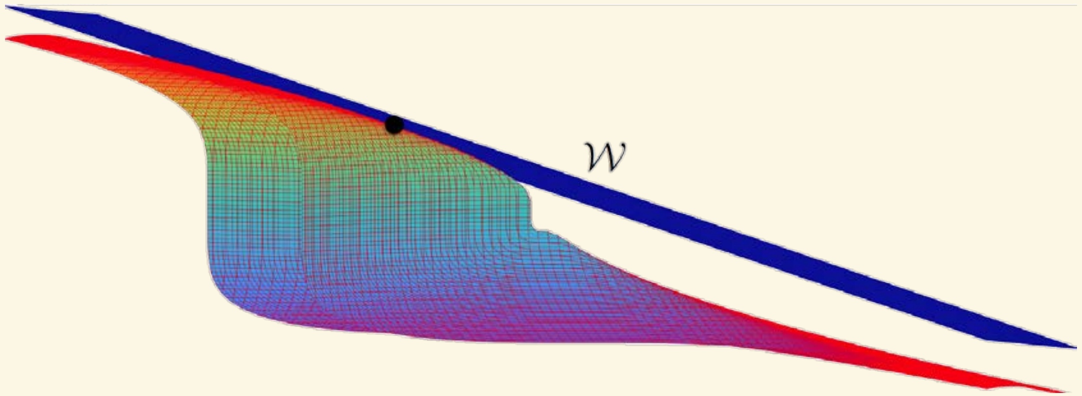
Non convex set!

Input: Linear subspace $W \subseteq \mathbb{C}^{d^2}$

Goal: Find rank one matrix ϵ-close to W

Top e-space of M: Linear subspace $W \subseteq \mathbb{C}^{d^2}$

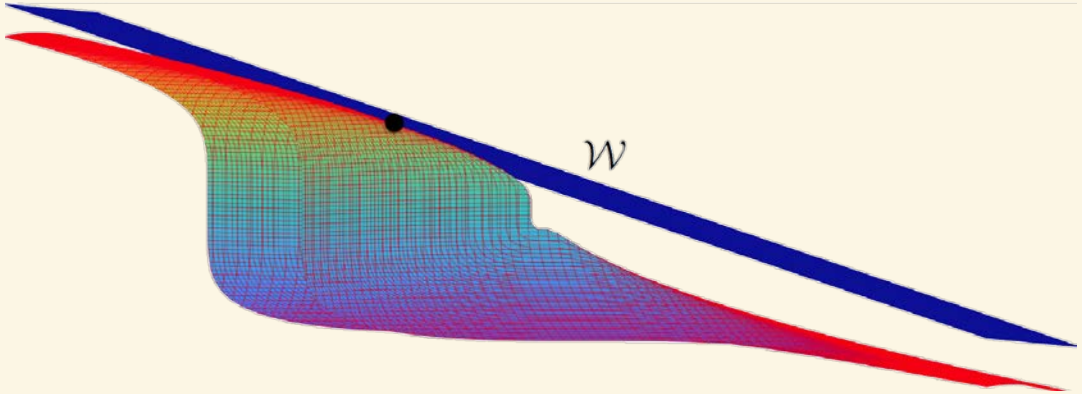
Goal: Find out if they intersect



Input: Linear subspace $W \subseteq \mathbb{C}^{d^2}$

(assume W contains rank one matrix)

Goal: Find rank one matrix ϵ-close to W



Trivial (brute force) **algorithm:** $2^{O(d)}$ time

Hardness: NP hard if $\epsilon = \frac{1}{poly(d)}$ [Gurvits'03, Gharbian'10]

Requires $d^{\Omega(\log d)}$ time for constant ϵ [Harrow-Montanaro'13]

$d^{O(\log d)}$ algorithm if M is 1-LOCC [Brandão-Christiandl-Yard'11]

Our Result: [B-Kothari-Steurer] $2^{\tilde{O}(\sqrt{d})}$ time

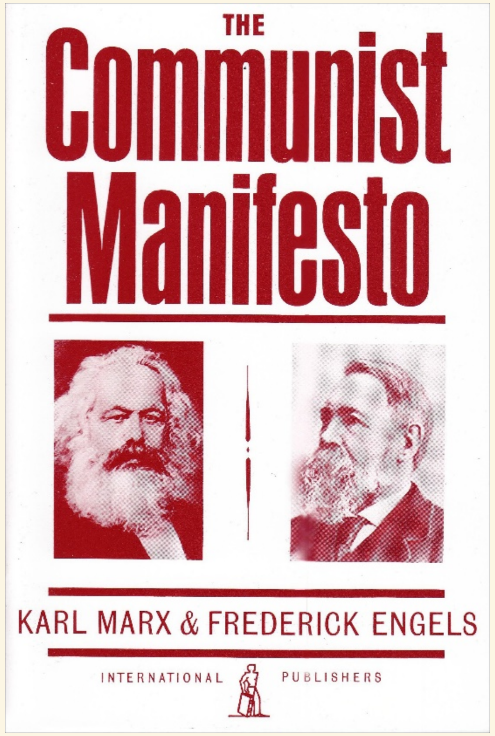
Analysis of algorithm of [Doherty-Parrilo-Spedalieri'04]

Sum of Squares

Sum of Squares Paradigm

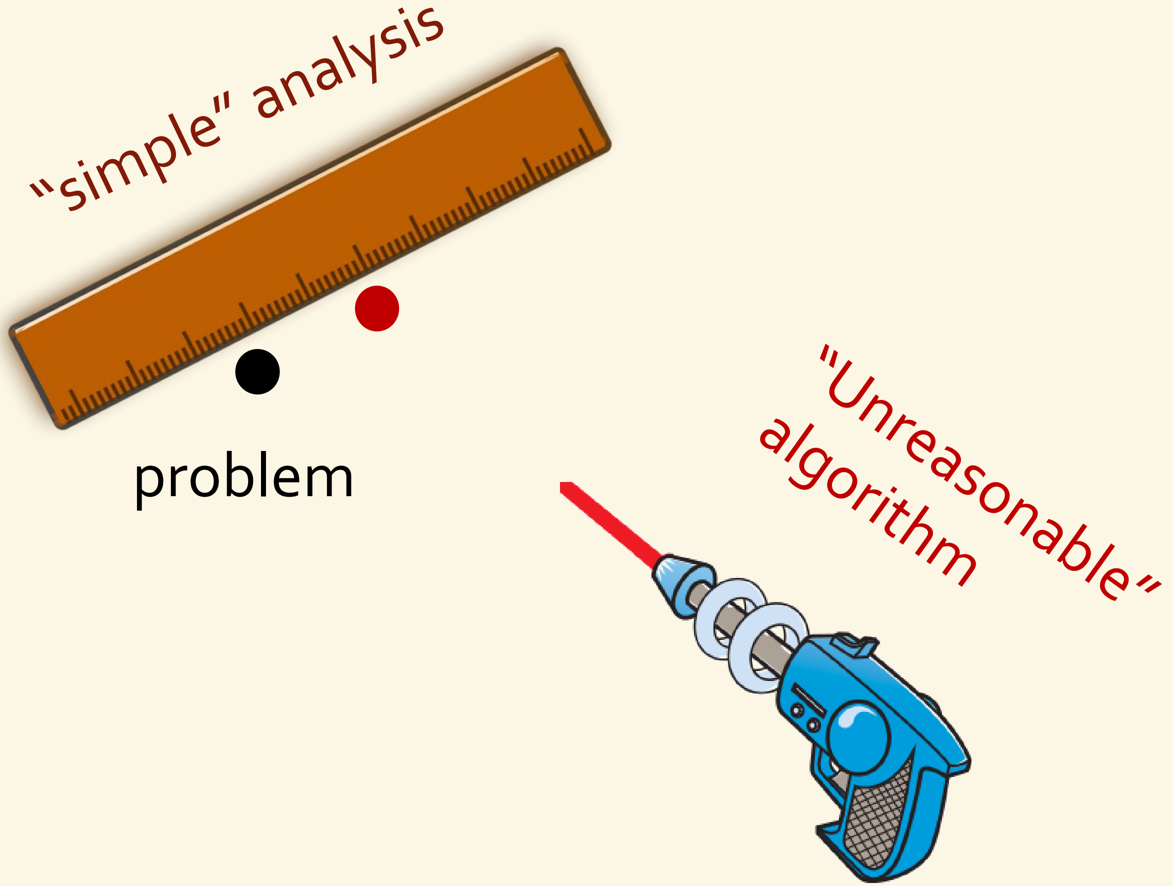
Observation: It's better to be rich than poor [Marx-Engels 1848]

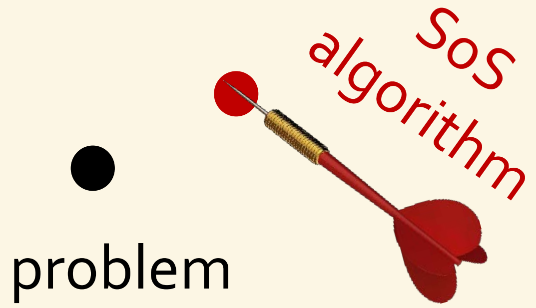
Easier to solve problems with extra power.



Sum of Squares Paradigm

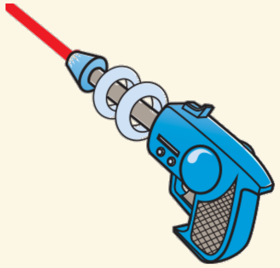
Observation: Easier to solve problems with extra power.





General philosophy

1) Prove correctness of "hypothetical" algorithm:



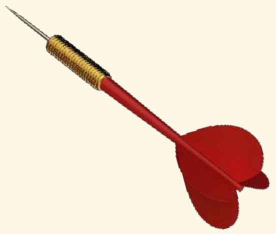
- *Has unbounded time.*

  Identifiability →recovery: Sparse coding, mixture models, community recovery, tensor completion, …

- *Gets "hints" about solution.*

  Combining →rounding: Sparse vector problem, best separable state

2) "Lift" proof to show SoS succeeds as well.



Deg ℓ SoS Proof System

[Artin'27,Krivine'61,Stengle'71,…,Grigoriev-Vorobjov'01]

AXIOM: $p^2 \geq 0$ for $\deg p \leq \ell$

Surprisingly powerful:

- Cauchy Schwarz
- Holder
- Hypercontractivity
- Invariance principle
- …

[B-Brandao-Harow-Kelner-Steurer-Zhou'12, De-Mossel-Neeman'12 ,O'Donnell-Zhou'13,Kauers-O'Donnell-Tan-Zhou'14,..]

Deg ℓ SoS Algorithm

[N.Shor'87,..,Parrilo'00,Lasserre'01]

INPUT: polynomial constraints on $x_1, \ldots, x_d$

OUTPUT: "fake" moments of distribution $\mathcal{D}$ over $\mathbb{R}^d$ satisfying constraints (via semi-definite programming)

SoS Proof System can't prove they are fake!

Sum of Squares and Quantum Information

Bell's Inequality: Alice gets $a \in \{0,1\}$ and outputs $X_a \in \{\pm 1\}$, Bob gets $b \in \{0,1\}$ and outputs $Y_b \in \{\pm 1\}$. Then

$$R = X_0Y_0 + X_0Y_1 + X_1Y_0 - X_1Y_1 \le 2$$

Pf: $R = X_0(Y_0 + Y_1) + X_1(Y_0 - Y_1) \le \sqrt{\left[X_0^2 + X_1^2\right] \cdot \left[\, (Y_0 + Y_1)^2 + (Y_0 - Y_1)^2 \,\right]}$

$$R \le \left| \sqrt{2\left(2Y_0^2 + 2Y_1^2\right)} \right| = \left|\sqrt{8}\right| \approx 2.828 \blacksquare$$

Quantum value of this game is $R = \sqrt{8}$!

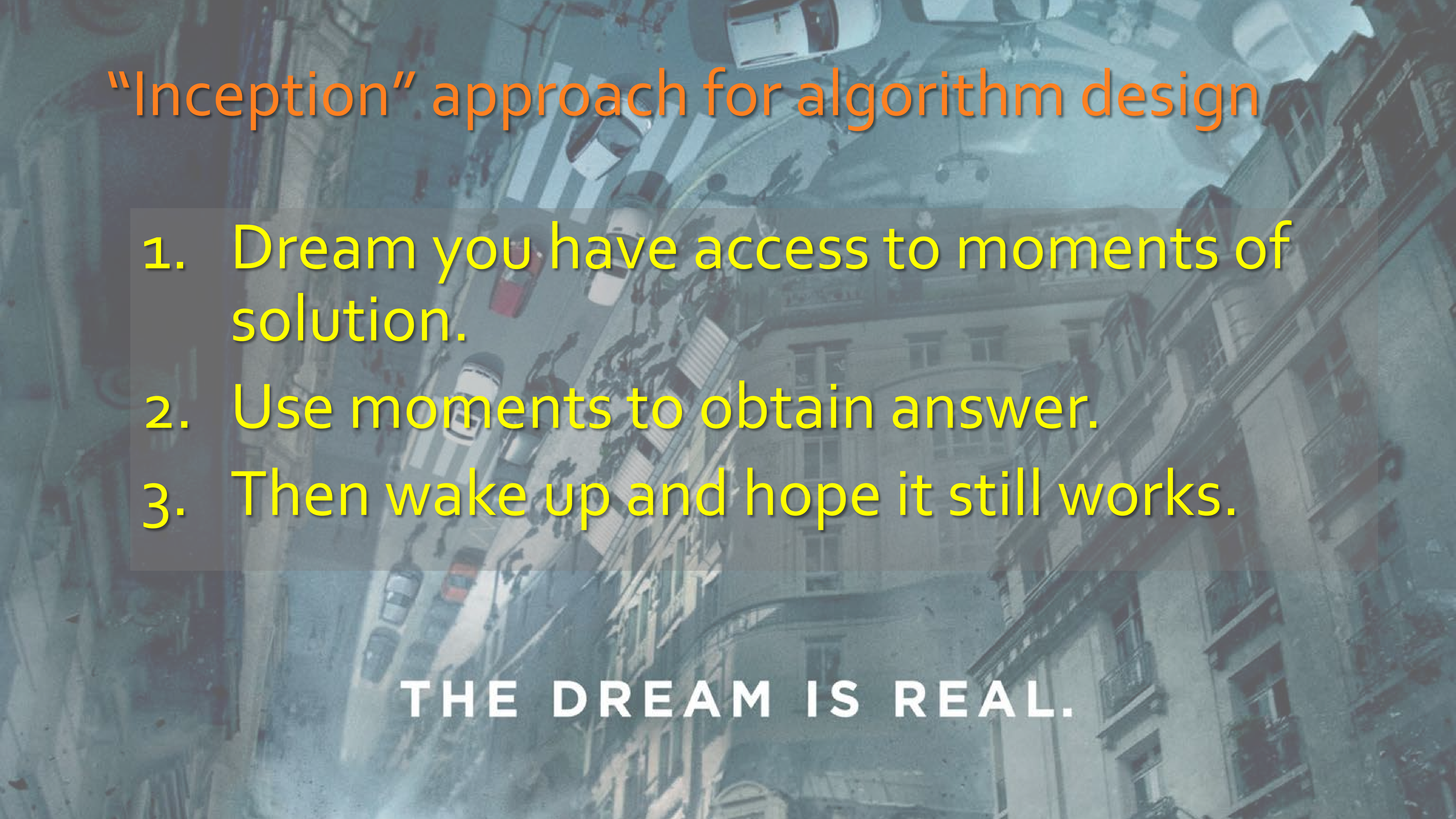
Nature might not follow Einstein..

.. but she does respect Cauchy-Schwarz

Solving Best Separable State via the Sum of Squares Algorithm

“Inception” approach for algorithm design

1. Dream you have access to moments of solution.
2. Use moments to obtain answer.
3. Then wake up and hope it still works.

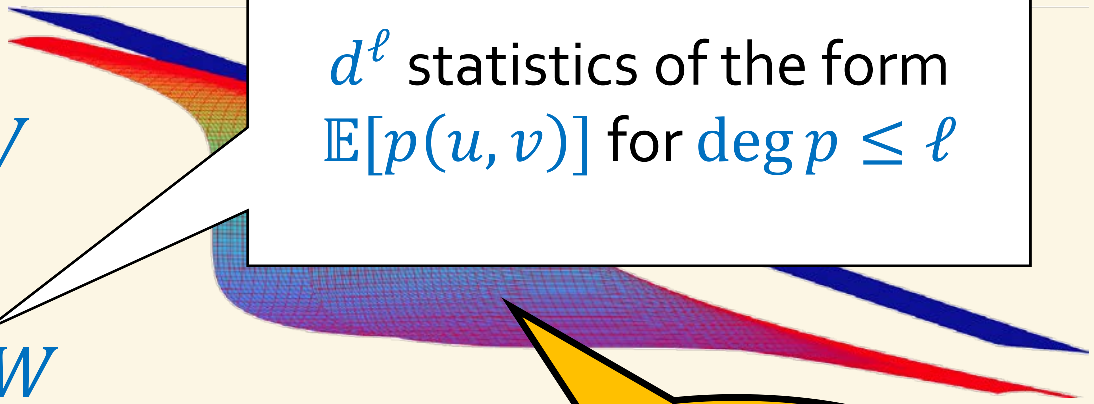


SoS Algorithm for Best Separable State

Input: Subspace $W \subseteq \mathbb{C}^{d^2}$

Assumption: Exists rank one matrix $|u\rangle\langle v| \in W$

Extra assumption: Get degree ℓ moments of distribution over $rank\ 1 \cap W$



d^ℓ statistics of the form $\mathbb{E}[p(u,v)]$ for $\deg p \leq \ell$

not convex!

Goal: Find rank one matrix $|\tilde{u}\rangle\langle\tilde{v}|$ ϵ-close to W

First Attempt: Can compute $\bar{A} = \mathbb{E}[|u\rangle\langle v|]$ (degree two moments $\mathbb{E}[u_i v_j]$)

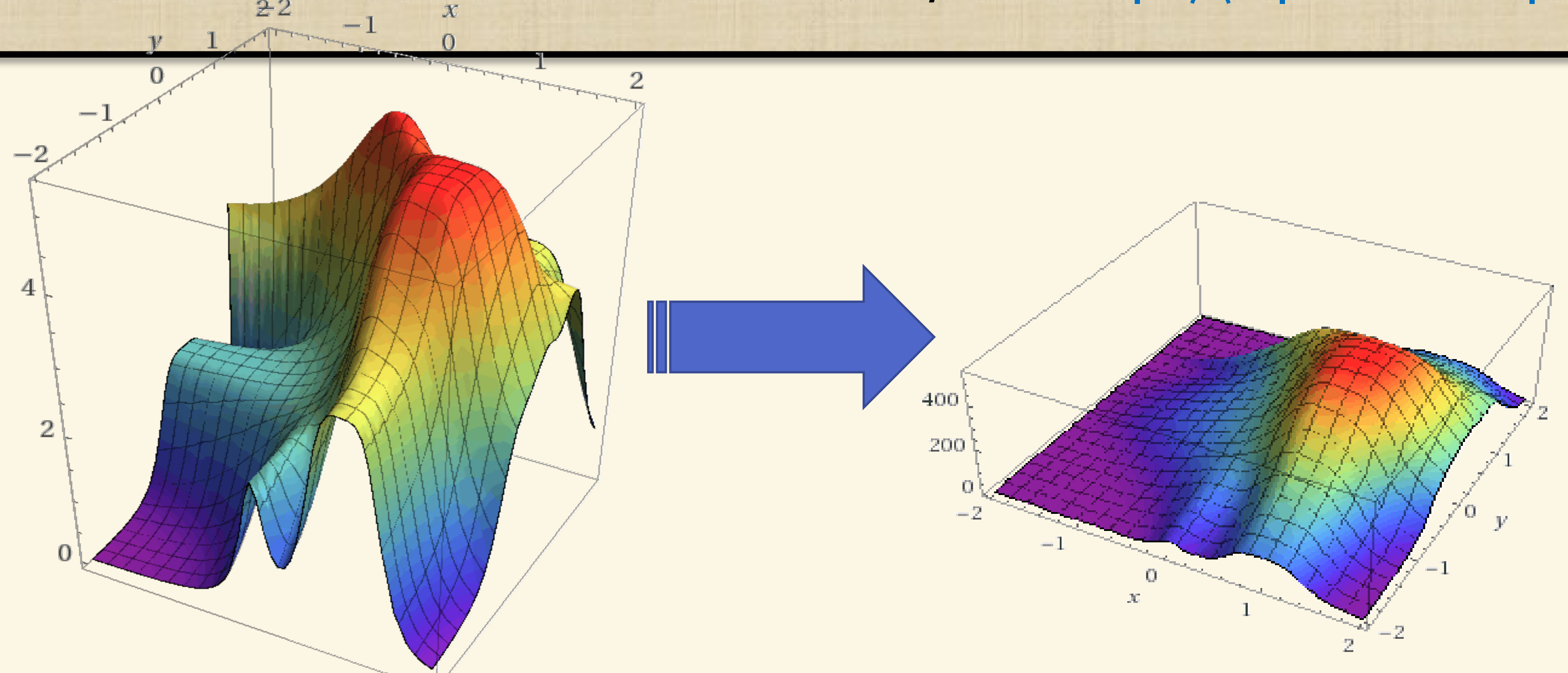
Dist over matrices in $W \Rightarrow \bar{A}$ is in W

$\bar{A}$ might not be rank one

“Convexifying” rank one matrices

THM: [B-Kothari-Steurer’17] For every dist $\mathcal{D}$ over rank one $d \times d$ matrices, exists $\tilde{O}(\sqrt{d})$ deg p s.t. $\overline{A} = \mathbb{E}[p(A)A]$ is “almost rank one”

(i.e., $\| \overline{A} - |u\rangle\langle v| \| \leq \epsilon \| |u\rangle\langle v| \|$)



"Convexifying" rank one matrices

THM: [B-Kothari-Steurer'17] For every dist $\mathcal{D}$ over rank one $d \times d$ matrices, exists $\tilde{O}(\sqrt{d})$ deg p s.t. $\overline{A} = \mathbb{E}[p(A)A]$ is "almost rank one"

(i.e., $\| \overline{A} - |u\rangle\langle v| \| \leq \epsilon \| |u\rangle\langle v| \|$)

Why would that be true? And where does $\sqrt{d}$ come from?

Inspiration: Lovett " sqrt rank theorem" in communication complexity

(p Schatten norm $\Rightarrow$ deg $\approx d^{1/p}$)

THM: [B-Kothari-Steurer'17] For every dist $\mathcal{D}$ over rank one $d \times d$ matrices,

exists $\tilde{O}(\sqrt{d})$ deg p s.t. $\overline{A} = \mathbb{E}[p(A)A]$ is "almost rank one"

(i.e., $\| \overline{A} - |u\rangle\langle v| \| \leq \epsilon \| |u\rangle\langle v| \|$)

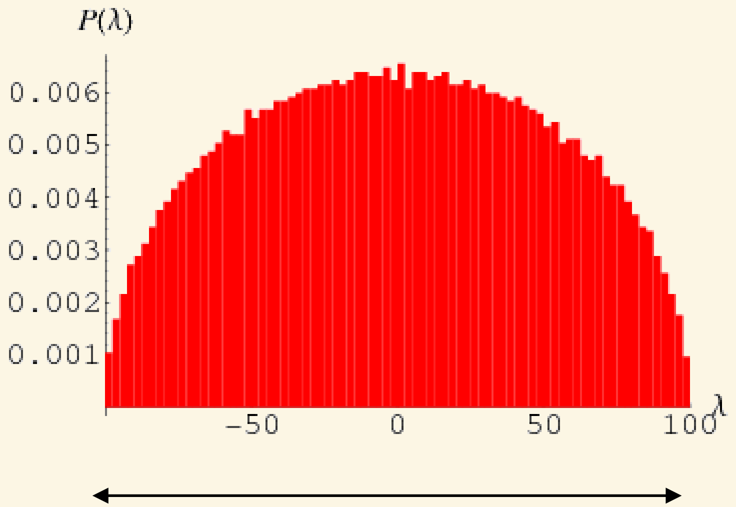
THM: [B-Kothari-Steurer'17] For every dist $\mathcal{D}$ over rank one $d \times d$ matrices,

exists $\tilde{O}(\sqrt{d})$ deg p s.t. $\bar{A} = \mathbb{E}[p(A)A]$ is "almost rank one"

(i.e., $\| \bar{A} - |u\rangle\langle v| \| \le \epsilon \cdot \| |u\rangle\langle v| \|$) (i.e., $\lambda_1^2 \gg \lambda_2^2 + \lambda_3^2 + \cdots + \lambda_d^2$)

Intuition: Random matrix eigenvalues follow Wigner semicircle law :

magnitude bounded by $O(\sqrt{d})$



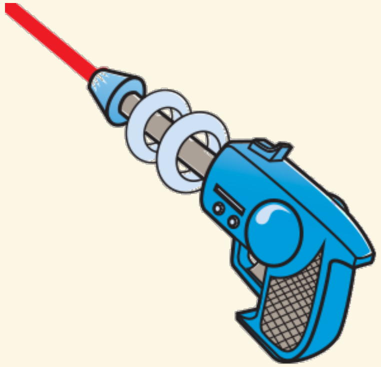
$-\sqrt{2d}$ $+\sqrt{2d}$

"Reweigh" dist by $\mathrm{p}(A) = (\langle u|A|u\rangle)^\ell$

$\Rightarrow$ Boosts e-val of u by $\frac{\mathbb{E}[N^{\ell+2}]}{\mathbb{E}[N^\ell]\mathbb{E}[N^2]} \approx \ell$

Get $\lambda_1 \approx \ell\sqrt{d} \gg d$ and $\lambda_2, \dots, \lambda_d \approx \sqrt{d}$

Recap

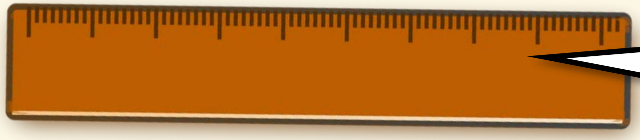


What we have

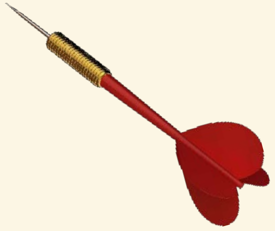
Given:

- W subspace of $\mathbb{C}^{d^2}$
- Degree $\tilde{O}(\sqrt{d})$ moments of $rank\ 1 \cap W$

Can find rank one matrix close to W



Proof of fact is in SOS framework!

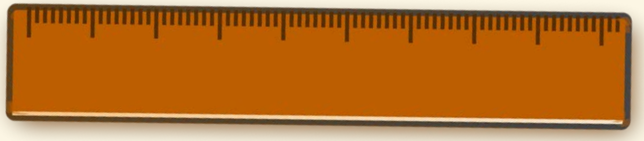


What we want

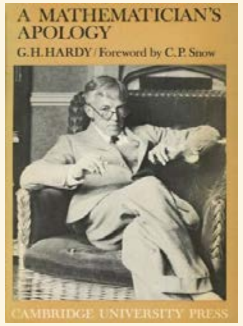
Given:

- W subspace of $\mathbb{C}^{d^2}$
- Nothing else
- “Fake” $\tilde{O}(\sqrt{d})$ moments generated via SoS

Find rank one matrix close to W

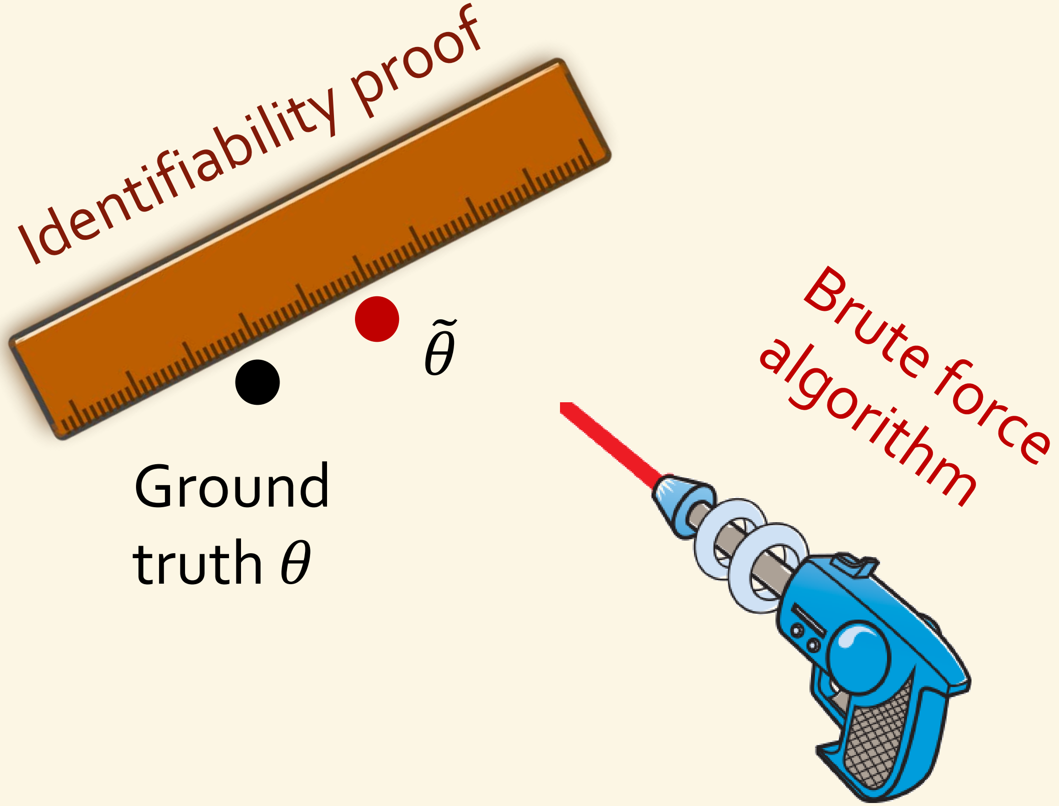


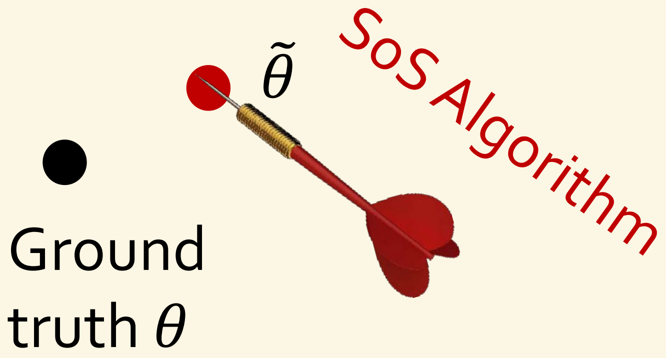
Other applications: unsupervised learning



Observations: $x_1, x_2, x_3, \ldots$ from model $P(\theta)$

Goal: Recover θ



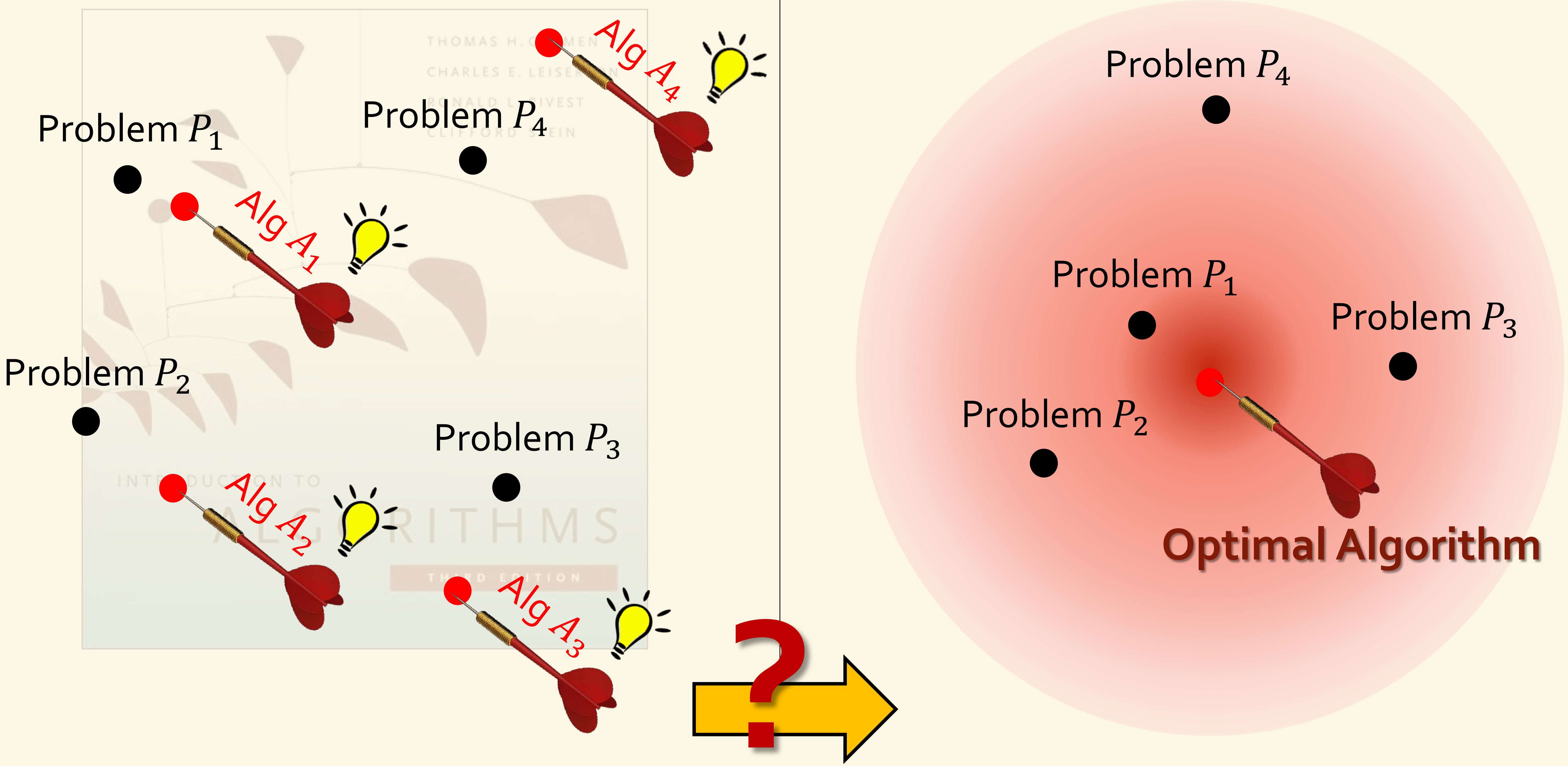


Example applications

- Dictionary learning [B-Kelner-Steurer'14,...]
- Tensor decomposition [Ge-Ma'15, Ma-Shi-Steurer'16]
- Tensor completion [B-Moitra'16, Potechin-Steurer'17]
- Tensor PCA [Hopkins-Shi-Steurer'15]
- Community detection [Hopkins-Steurer'17]
- Gaussian Mixture Models [Hopkins-Lin'17, Kothari-Steinhardt'17]
- Outlier-robust estimation [Kothari-Steurer'17]

THE DREAM IS REAL.

A different approach to algorithm design



Summary

- Better than brute force ($2^{\tilde{O}(\sqrt{d})}$ time) alg for best separable state
- Still large gap from $2^{\Omega(\log d)}$ lower bound.
- Still open: noisy version, quantum separability problem
- Ideas lead to improved algorithms in several worst case and average case settings.
- Often SoS based algorithm gives best known guarantees.
- Is this accidental? Or part of larger pattern?

Thank you!