The Power of Sum of Squares for Detecting Hidden Structures

Tselil Schramm
(Harvard & MIT)

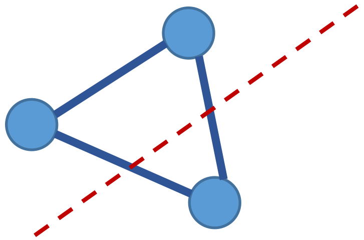
with Sam Hopkins (Cornell University), Pravesh Kothari (Princeton/IAS),
Aaron Potechin (KTH), Prasad Raghavendra (UC Berkeley),
and David Steurer (ETH Zurich).

Sum-of-Squares $\equiv_{\text{avg}}$ Spectral Algorithms

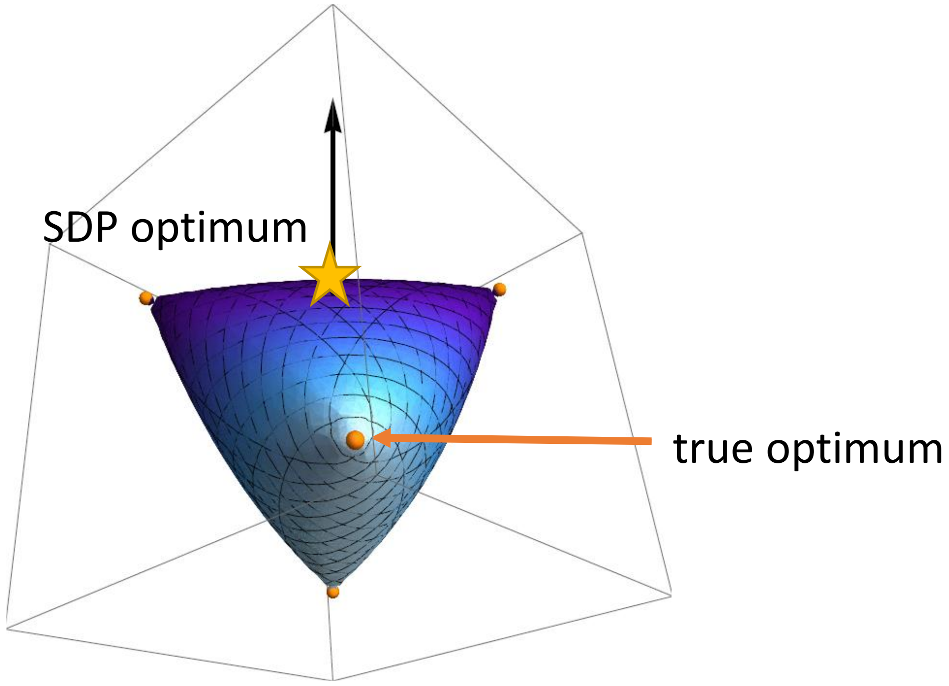
Tselil Schramm
(Harvard & MIT)

with Sam Hopkins (Cornell University), Pravesh Kothari (Princeton/IAS),
Aaron Potechin (KTH), Prasad Raghavendra (UC Berkeley),
and David Steurer (ETH Zurich).

Semidefinite programs

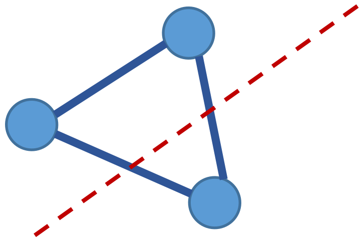


e.g. max-cut

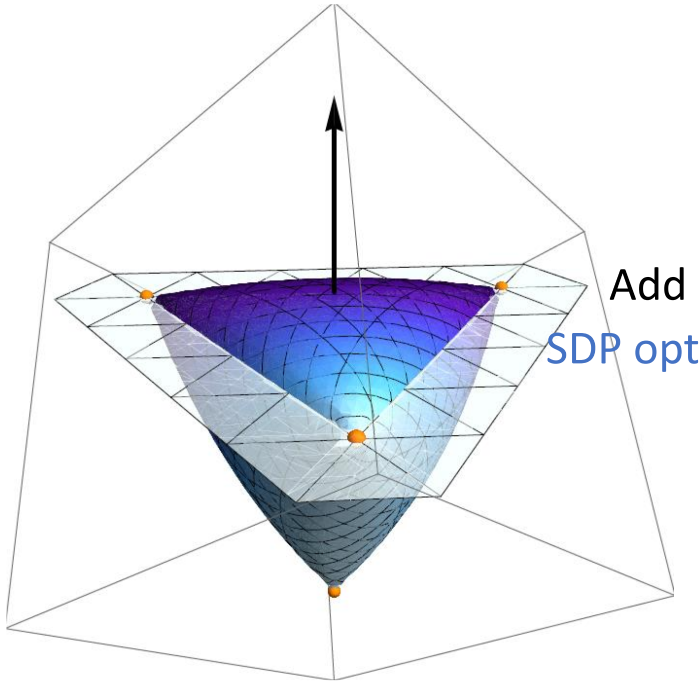


Non-convex problem → SDP Relaxation

Semidefinite programs



e.g. max-cut

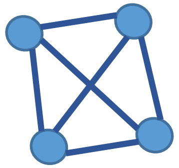


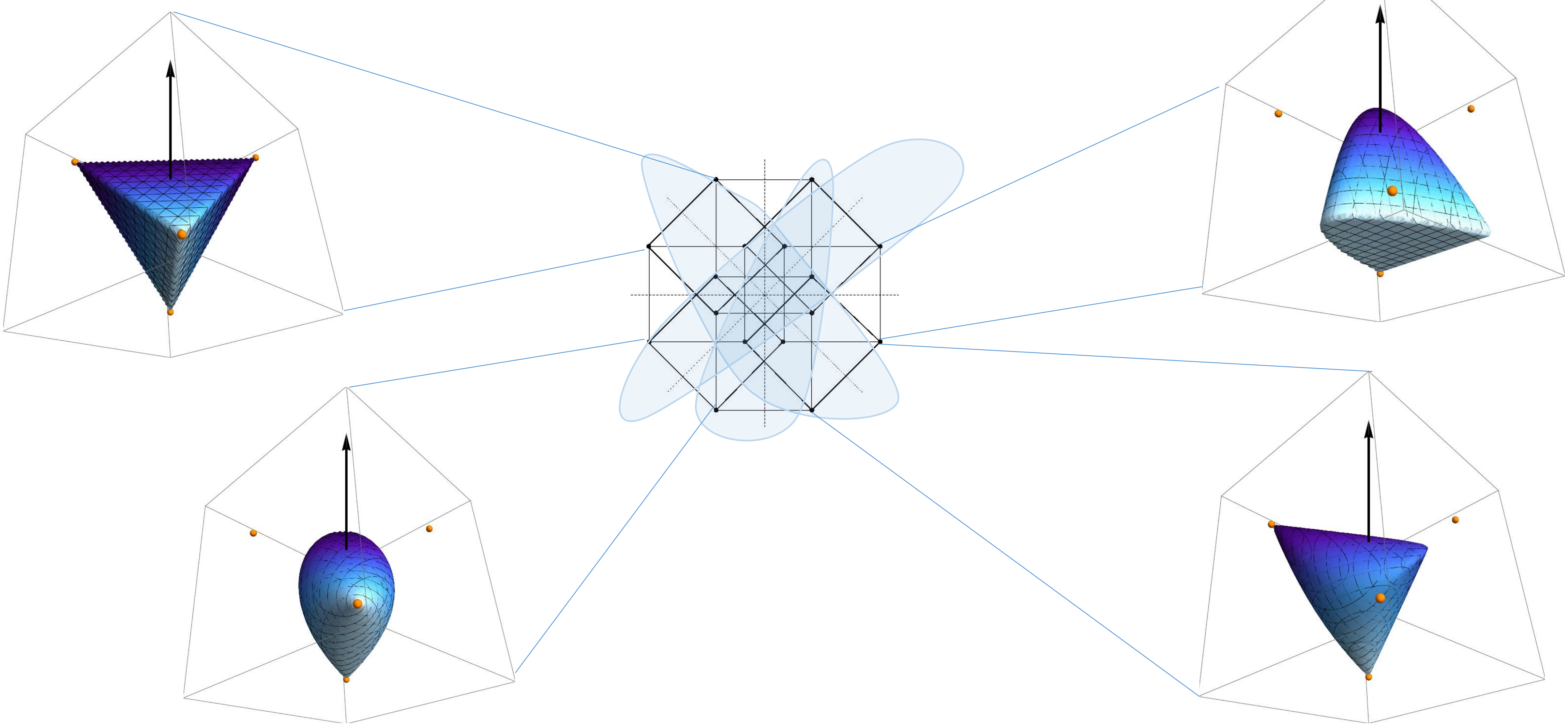
Add a new constraint!

SDP optimum = true optimum

Non-convex problem → SDP Relaxation

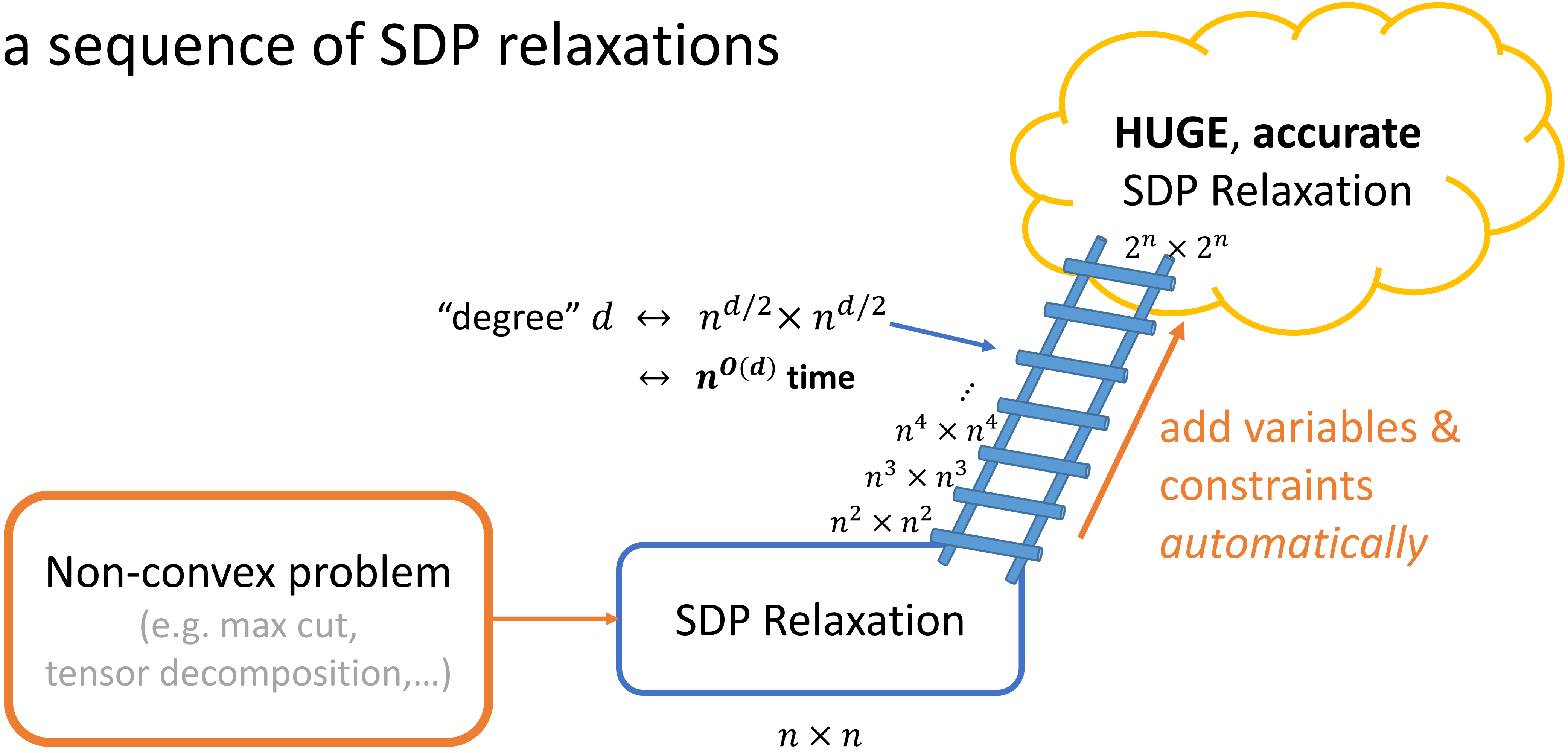
How to find tighter SDPs?





The Sum-of-Squares Hierarchy.

a sequence of SDP relaxations



Non-convex problem
(e.g. max cut,
tensor decomposition,...)

SDP Relaxation

$n \times n$

$n^2 \times n^2$

$n^3 \times n^3$

$n^4 \times n^4$

$\therefore$

"degree" d ↔ $n^{d/2} \times n^{d/2}$

↔ $\boldsymbol{n^{O(d)}}$ **time**

HUGE, accurate
SDP Relaxation
$2^n \times 2^n$

add variables &
constraints
automatically

Why study Sum-of-Squares (SoS)?

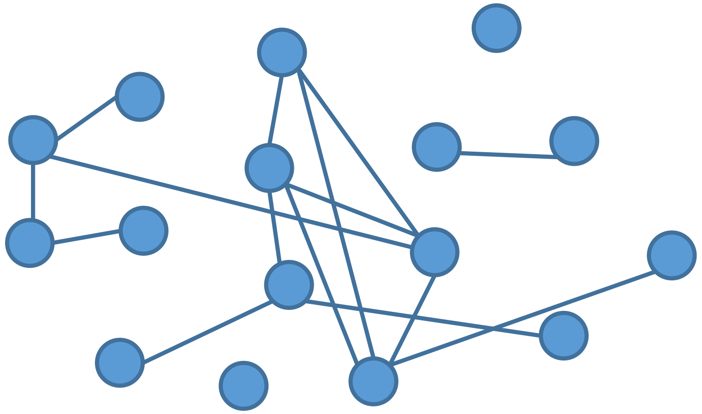
- Principled way to tighten SDPs.
- Expressive Power.
  SoS is the *tightest* SDP for many problems [Lee-Raghavendra-Steurer ‘15]
- Disprove Unique Games Conjecture? [Barak-Brandão-Harrow-Kelner-Steurer-Zhou ‘12]
- New algorithms for Average-Case Problems!

Why study Sum-of-Squares (SoS)?

- Principled way to tighten SDPs.
- Expressive Power.
  SoS is the *tightest* SDP for many problems [Lee-Raghavendra-Steurer ‘15]
- Disprove Unique Games Conjecture? [Barak-Brandão-Harrow-Kelner-Steurer-Zhou ‘12]
- New algorithms for Average-Case Problems!

Average-Case Problems

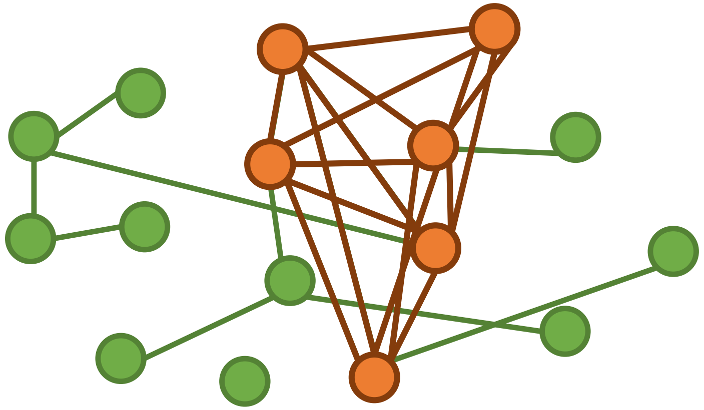
e.g. Planted Clique



random graph

“null hypothesis”

vs.



random graph + k-clique

structure/large objective value

Average-Case Problems

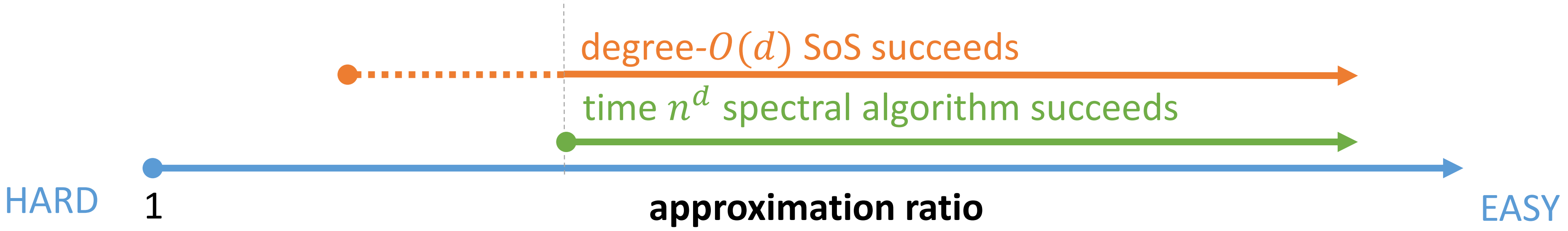
SoS semidefinite program + rounding gives new algorithms for average-case problems.

- Refuting Random CSPs [Allen-O'Donnell-Witmer'15, Raghavendra-Rao-**S**'17]
- Tensor Decomposition/Dictionary Learning [Barak-Kelner-Steurer'14, Ge-Ma'15, Ma-Shi-Steurer'16]
- Planted Sparse Vector [Barak-Brandão-Harrow-Kelner-Steurer-Zhou'12, Barak-Kelner-Steurer'14]
- Tensor Completion [Barak-Moitra'16, Potechin-Steurer'17]
- Tensor PCA [Hopkins-Shi-Steurer'15,Bhattiprolu-Guruswami-Lee'16, Raghavendra-Rao-**S**'17]

Lower signal-to-noise ratios, looser assumptions about models, unified approach.

In the worst case, we think SoS ≥ Spectral

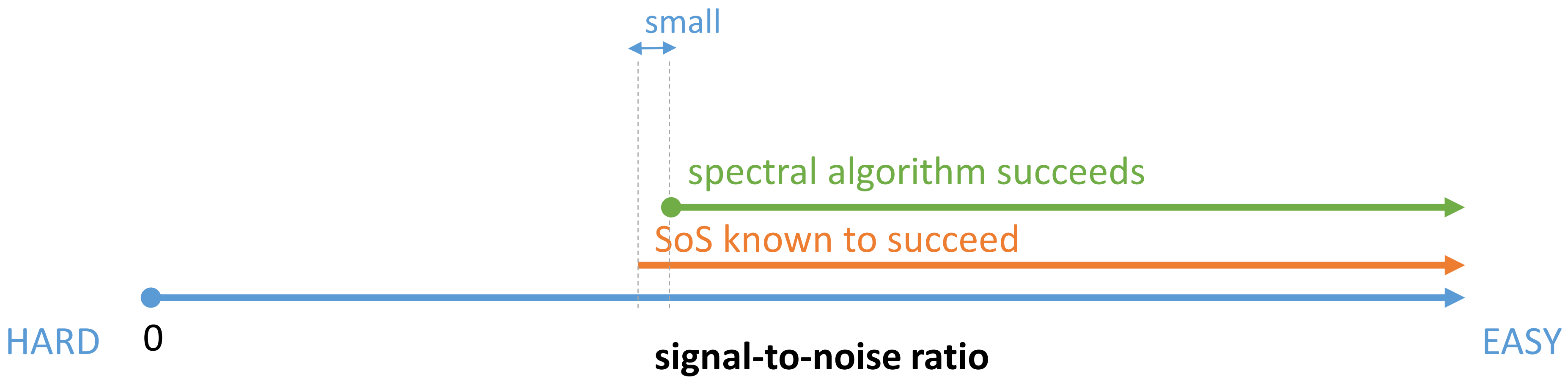
SoS captures spectral algorithms.



“Containment” sometimes strict (as far as we know), e.g.

	spectral approx	SoS approx
coloring 3-colorable graphs	?	$n^{0.2-\varepsilon}$ [Kawarabayashi-Thorup ‘17]
MAX-CUT	0.61 [Soto ‘14]	0.87 [Goemans-Williamson ‘94]

But intriguingly…



small

spectral algorithm succeeds

SoS known to succeed

HARD

0

signal-to-noise ratio

EASY

Often, if SoS succeeds for average case, there is also a possibly fancy spectral algorithm which succeeds.

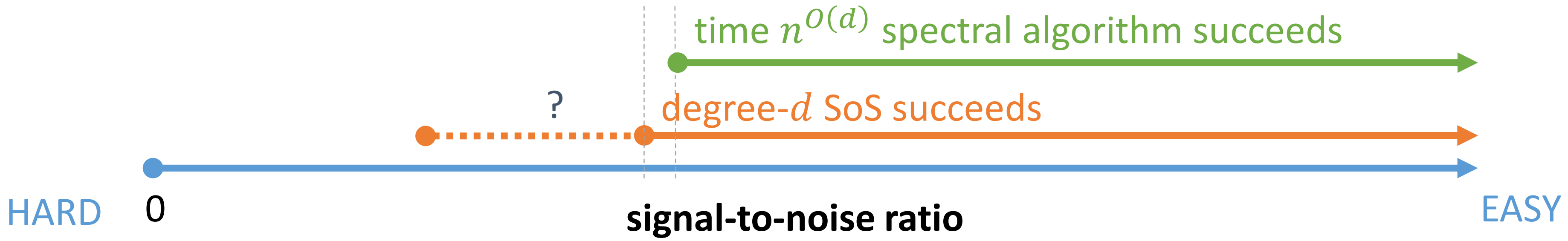
Refuting Random CSPs [Allen-O'Donnell-Witmer '15, Raghavendra-Rao-**S**'17]

Tensor PCA [Hopkins-Shi-Steurer'15, Bhattiprolu-Guruswami-Lee'16, Raghavendra-Rao-**S**'17]

Tensor Decomposition [Hopkins-**S**-Shi-Steurer'16, **S**-Steurer'17]

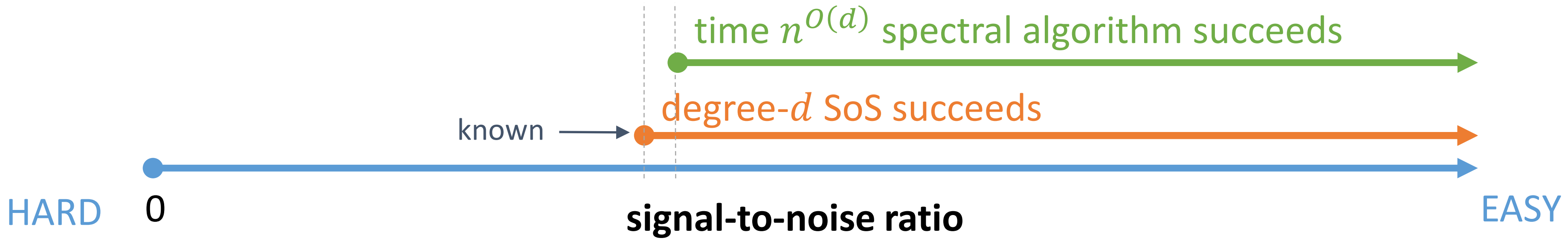
Tensor Completion [Montanari-Sun'17]

Is this more than anecdotal?



Is this more than anecdotal?

For many *average-case* problems, we *know* SoS $\lesssim$ spectral algorithms



Refuting Random CSPs [Allen-O'Donnell-Witmer '15, Raghavendra-Rao-**S**'17, Kothari-O'Donnell-Mori-Witmer'17]

Planted Clique [Alon-Krivelevich-Sudakov'98, Barak-Hopkins-Kelner-Kothari-Moitra-Potechin'16]

Tensor PCA [Hopkins-Shi-Steurer'15, Bhattiprolu-Guruswami-Lee'16, Raghavendra-Rao-**S**'17, **this paper**]

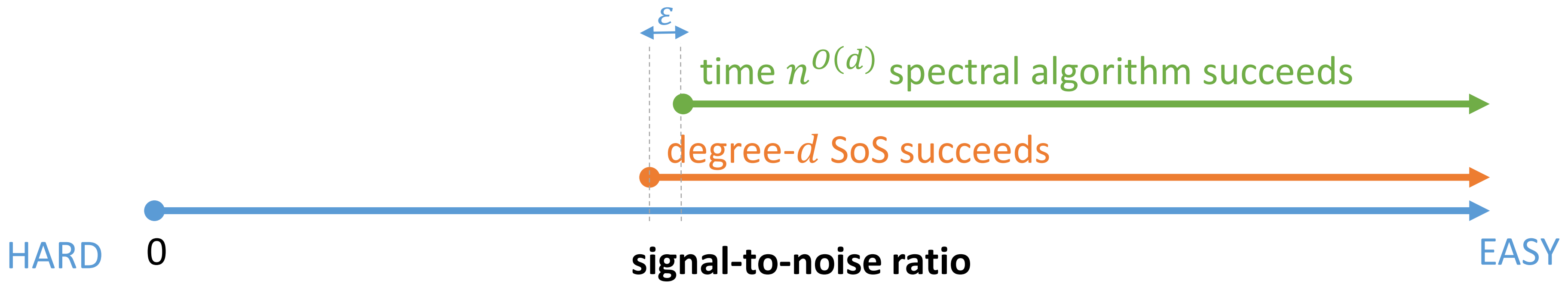
Sparse PCA [Deshpande-Montanari'14, Ma-Wigderson'15, **this paper**]

Result (quick version)

Let’s make these precise…

Theorem

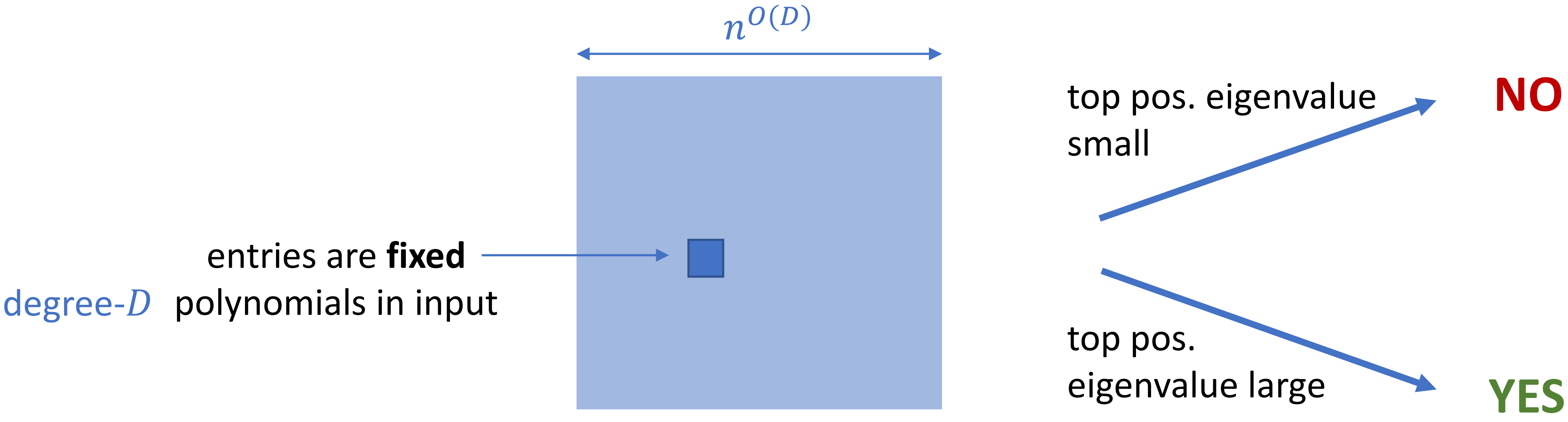
SoS and spectral algorithms are equivalent for average-case problems.



average-case problems: planted clique, refuting random CSPs, community detection/stochastic block models, densest-k-subgraph, tensor PCA, sparse PCA, …

Our notion of “spectral algorithm”

decision problem P



$n^{O(D)}$

entries are **fixed** degree-D polynomials in input

top pos. eigenvalue small → **NO**

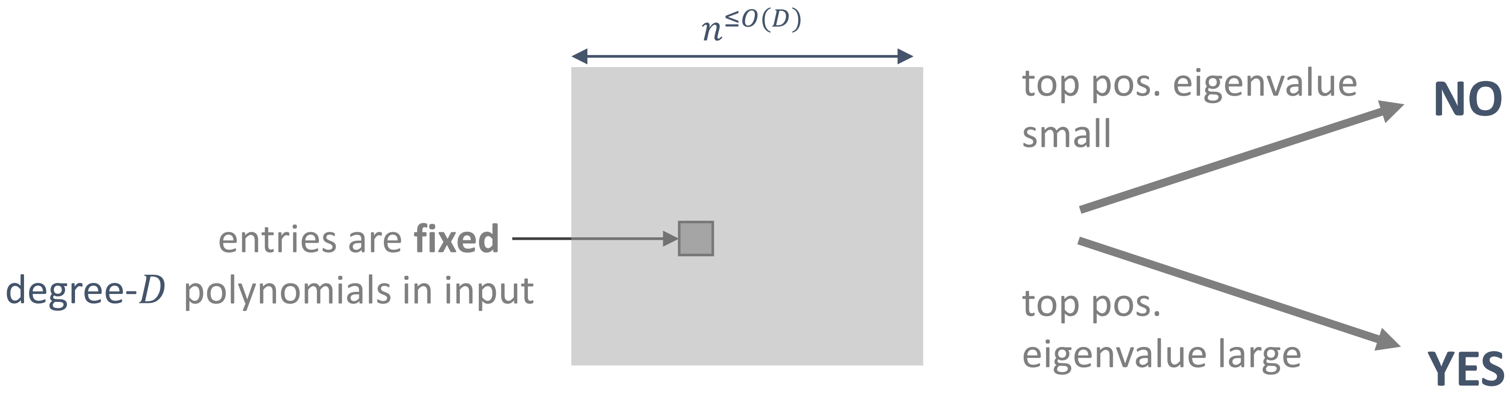
top pos. eigenvalue large → **YES**

degree-D spectral algorithm requires time $n^{O(D)}$ to construct matrix and compute top eigenvalue

For example...

decision problem P

e.g. P = does G contain clique of size $\geq k$?



$n^{\leq O(D)}$

entries are **fixed** degree-D polynomials in input

top pos. eigenvalue small

NO

top pos. eigenvalue large

YES

---

input $G = \{G_{ij}\}_{i,j\in[n]}$

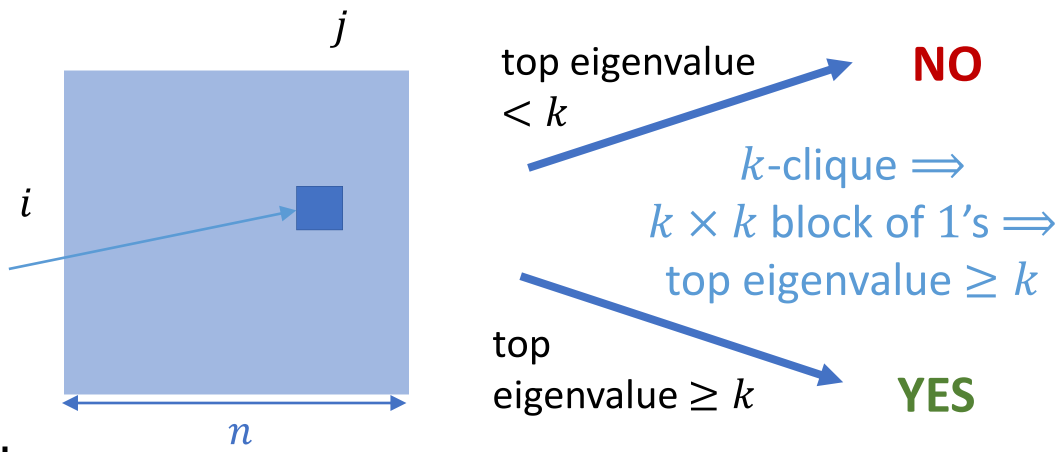
0/1 indicator for edge (i,j)

$$\boxed{M(G) = 2 \cdot A(G) - \vec{1}\vec{1}^{\top}}$$

degree D = 1

$$M_{ij}(G) = 2G_{ij} - 1$$

$$= \begin{cases} 1 & i \sim j \\ -1 & \text{o.w.} \end{cases}$$



j

i

n

top eigenvalue $< k$

NO

k-clique $\Longrightarrow$

$k \times k$ block of 1's $\Longrightarrow$

top eigenvalue $\geq k$

top eigenvalue $\geq k$

YES

completeness = 1

soundness bad in worst case,

nontrivial in average case

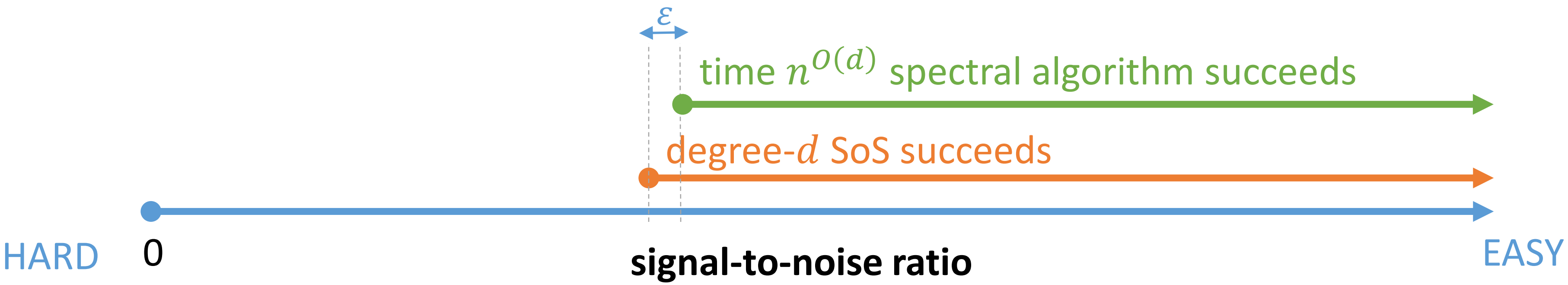
Result (quick version)

Let's make these precise…

Theorem

✓

SoS and spectral algorithms are equivalent for average-case problems.

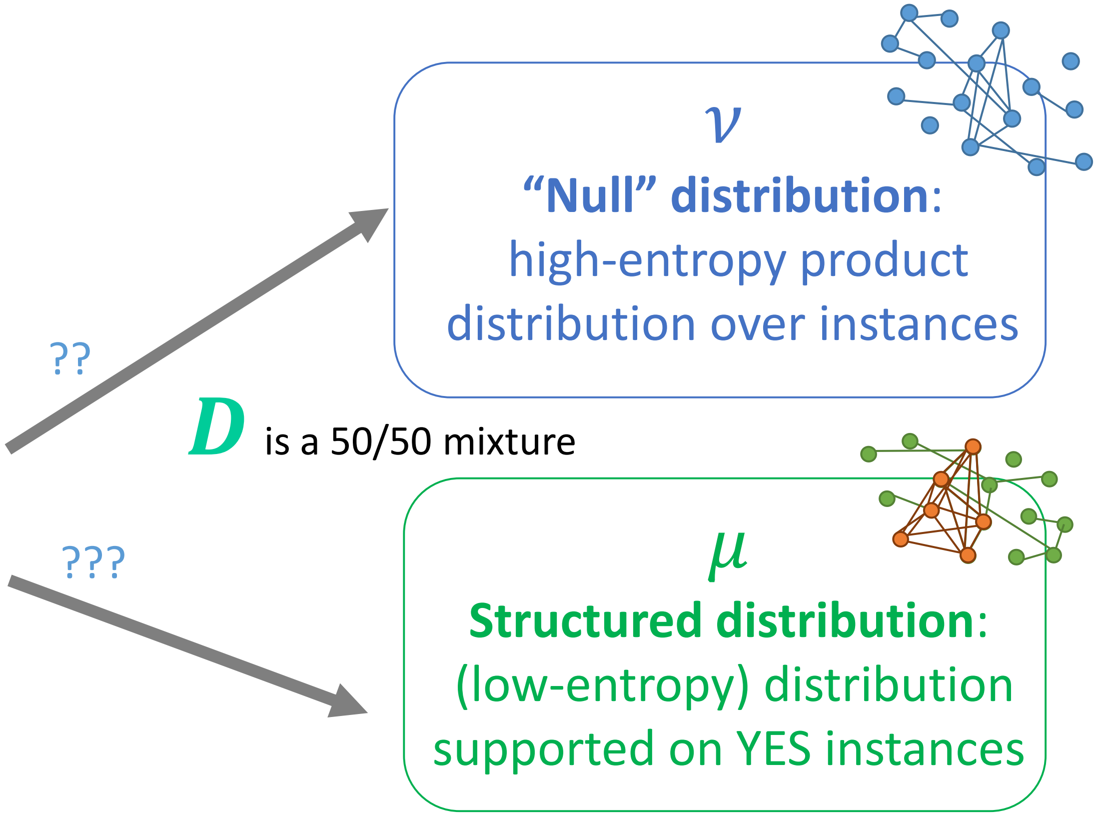


Average-Case Problems

decision problem P

e.g. does G have a clique of size $\geq k$?

Given $G \sim D$,
decide w/prob $\geq 1 - o(1)$
if $G \sim \mu$ or $G \sim \nu$



??

???

D is a 50/50 mixture

ν
“Null” distribution:
high-entropy product distribution over instances

μ
Structured distribution:
(low-entropy) distribution supported on YES instances

---

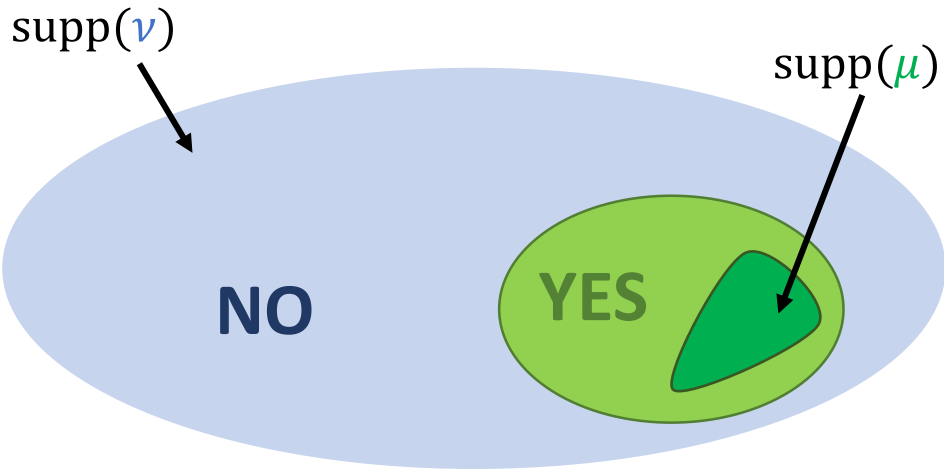
e.g. Planted Clique:

ν uniform over graphs $G\left(n, \frac{1}{2}\right)$

μ uniform over graphs containing clique of size k

SoS: w/prob > $1 - o(1)$ over $G \sim \nu$, SoS SDP value $\leq k$
if $G \sim \mu$, SoS SDP value $\geq k$

Spectral: w/prob > $1 - o(1)$ over $G \sim \nu$, eigenvalue is small
if $G \sim \mu$, eigenvalue large



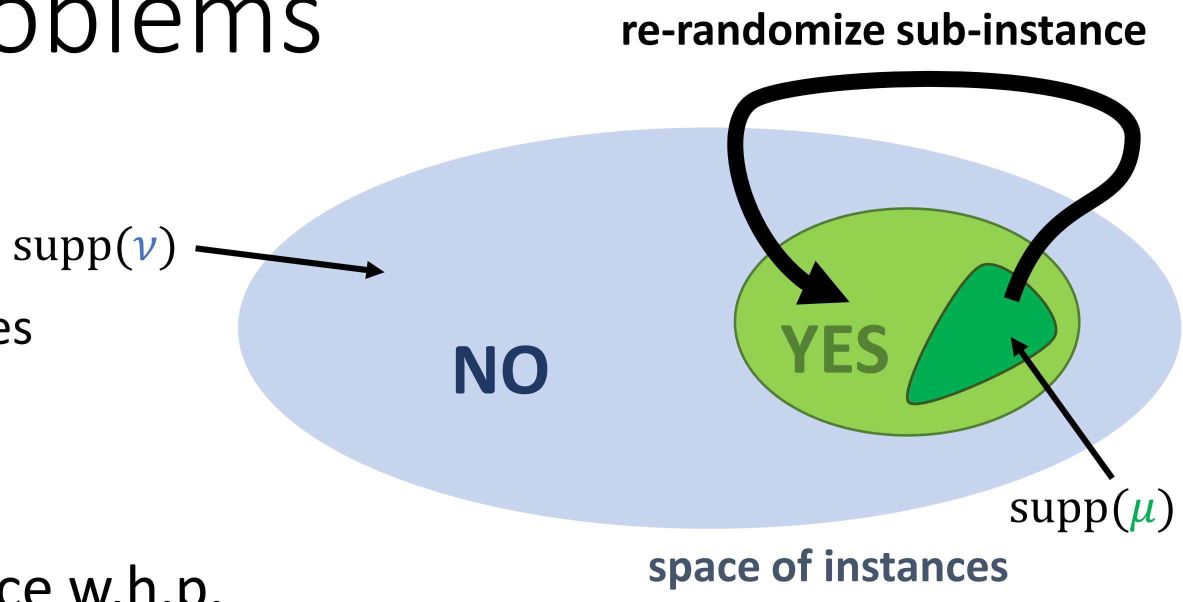
space of instances

Robust Average-Case Problems

decision problem P

structured distribution μ supported on YES instances

null distribution ν (NO instance w.h.p.)



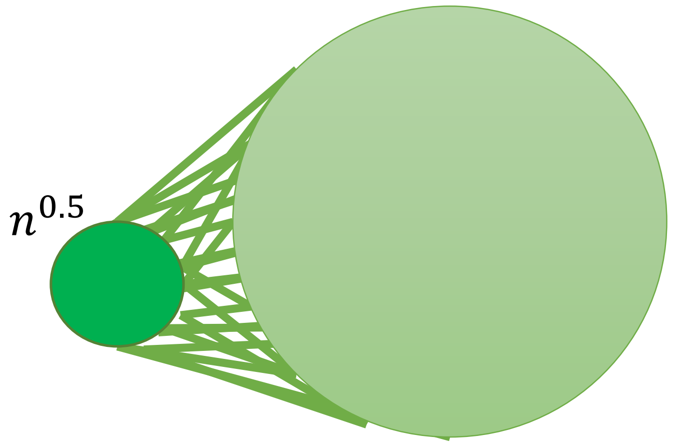
Problem is *robust* if $G \sim \mu$ is still a **YES** instance w.h.p. after re-randomizing some of the instance

---

e.g. P = G has a clique of size $\geq n^{0.4}$?

ν is $G\left(n,\frac{1}{2}\right)$

μ is $G\left(n,\frac{1}{2}\right) + n^{0.5}$-clique



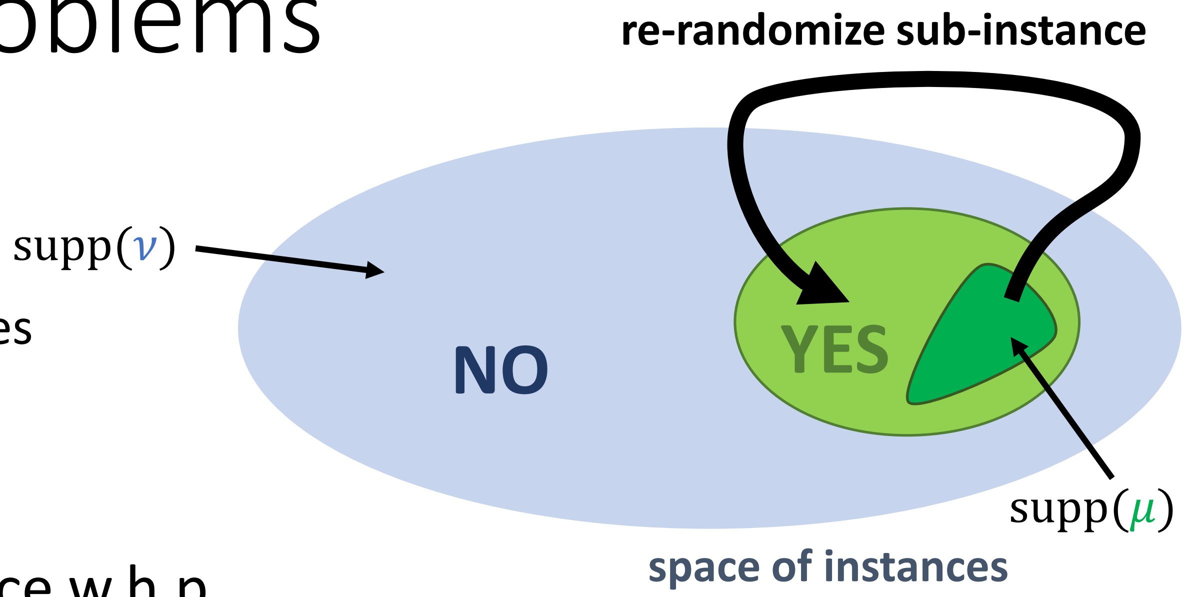
$\sim \mu$

Robust Average-Case Problems

decision problem P

structured distribution μ supported on YES instances

null distribution ν (NO instance w.h.p.)



re-randomize sub-instance

$\mathrm{supp}(\nu)$

NO

YES

$\mathrm{supp}(\mu)$

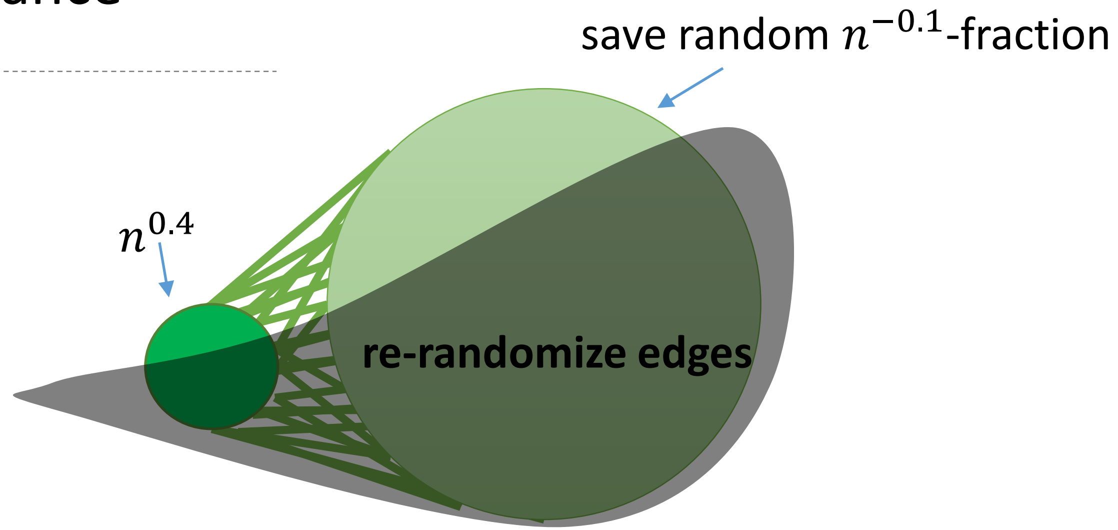
space of instances

Problem is *robust* if $G \sim \mu$ is still a **YES** instance w.h.p. after re-randomizing some of the instance

e.g. P = G has a clique of size $\geq n^{0.4}$?

ν is $G\left(n, \frac{1}{2}\right)$

μ is $G\left(n, \frac{1}{2}\right) + n^{0.5}$-clique



save random $n^{-0.1}$-fraction

$n^{0.4}$

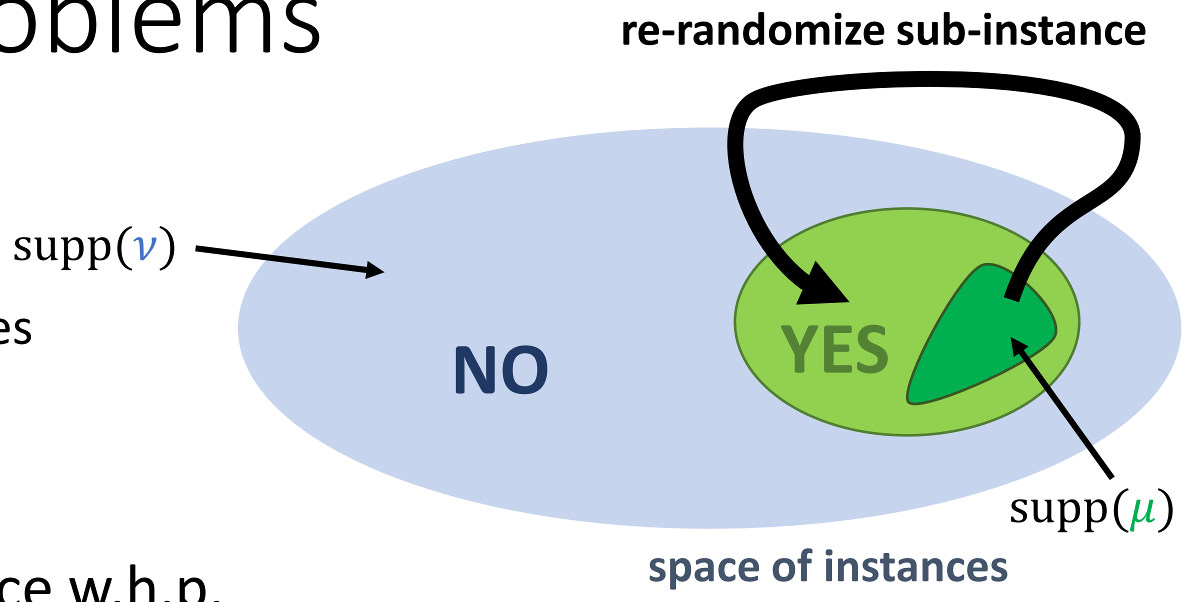
re-randomize edges

Robust Average-Case Problems

decision problem P

structured distribution μ supported on YES instances

null distribution ν (NO instance w.h.p.)



re-randomize sub-instance

supp(ν)

NO

YES

supp(μ)

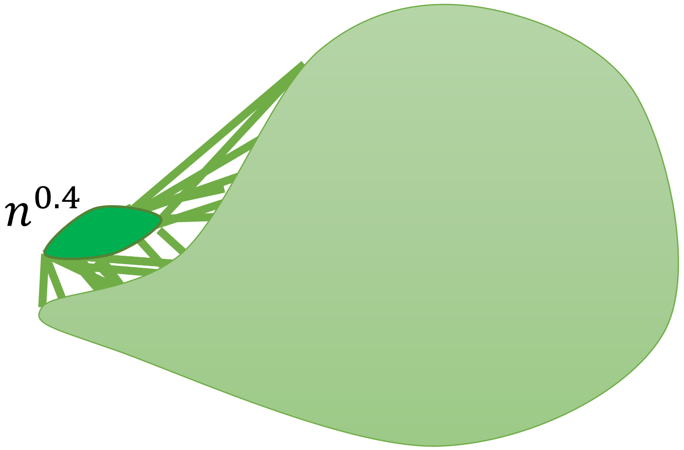
space of instances

Problem is *robust* if $G \sim \mu$ is still a **YES** instance w.h.p. after re-randomizing some of the instance

e.g. P = G has a clique of size $\geq n^{0.4}$?

ν is $G\left(n, \frac{1}{2}\right)$

μ is $G\left(n, \frac{1}{2}\right) + n^{0.5}$-clique

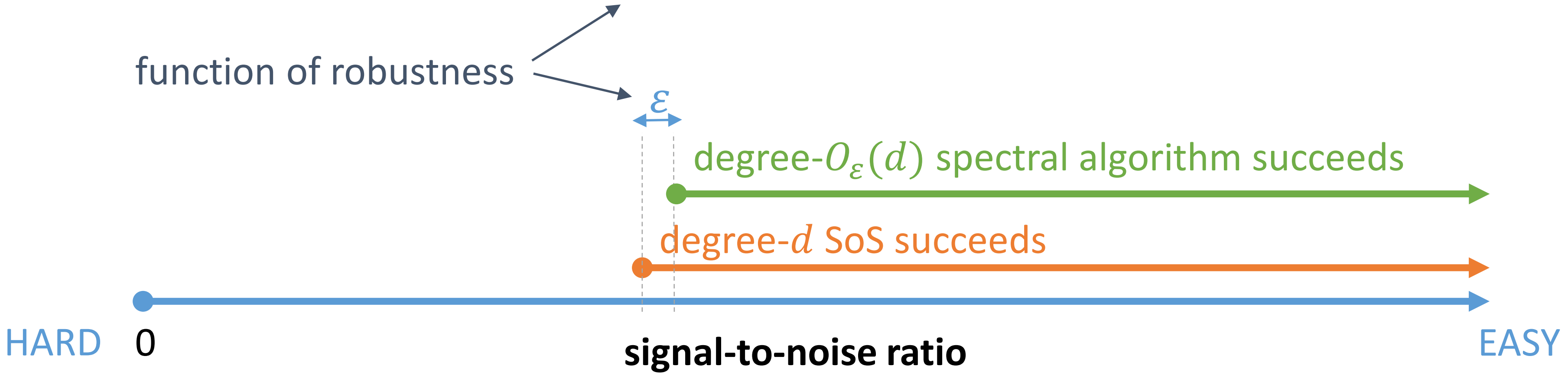


YES w.h.p.

Result (full(er) version)

Theorem

For robust average-case problems,
if there is a degree-d SoS algorithm then there is a
degree-$O_\varepsilon(d)$ spectral algorithm.



function of robustness

ε

degree-$O_\varepsilon(d)$ spectral algorithm succeeds

degree-d SoS succeeds

HARD 0

signal-to-noise ratio

EASY

Proof:

SoS thinks $OBJ \geq k$ w.h.p.
over null distribution

Want to show: if there is no spectral algorithm, then there is no SoS Algorithm.

Proof: by duality!

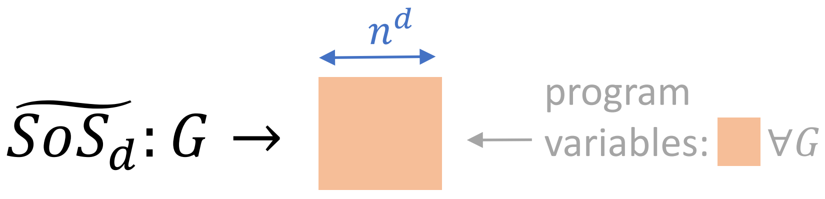
SoS thinks $OBJ \geq k$ w.h.p.
over null distribution

Want to show: if there is no spectral algorithm, then there is no SoS Algorithm.

Use duality with an **exponentially large** convex program.

Primal: solve for approximate SoS solution

$\min$ 🟧 $\forall G$ $\quad \mathrm{P}_{G\sim\nu}[OBJ($🟧$) \leq k]$ (-ish)



$\widetilde{SoS}_d : G \to$ 🟧 ⟵ program variables: 🟧 $\forall G$

s.t. " 🟧 satisfies SoS constraints on average"

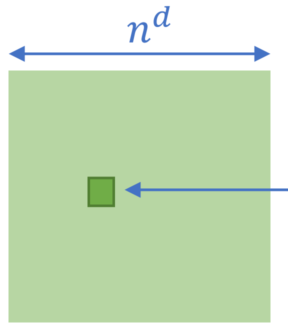
≈ **strong** ⟺ **duality**

Dual: solve for best spectral algorithm

$\max$ 🟩 $\quad \mathbb{E}_{G\sim\mu}[\lambda_{\max}($🟩$)]$ **YES**

s.t. $\mathbb{E}_{G\sim\nu}[\lambda_{\max}($🟩$)] \leq 1$ **NO**

where



entries are degree-D polynomials in G

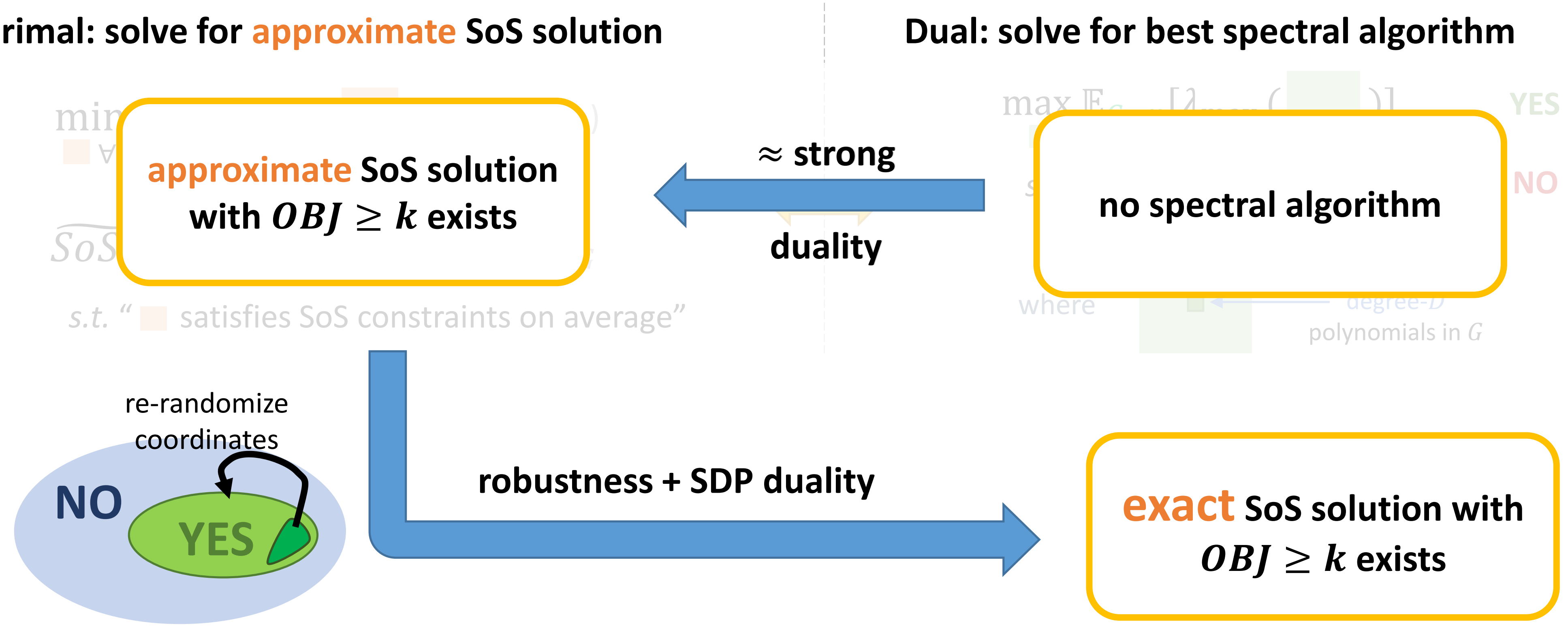
inspired by Barak et al. '16 planted clique lower bound

Proof: by duality!

SoS thinks $OBJ \geq k$ w.h.p.
over null distribution

Want to show: if there is no spectral algorithm, then there is no SoS Algorithm.

Use duality with an **exponentially large** convex program.



Primal: solve for approximate SoS solution

Dual: solve for best spectral algorithm

s.t. “ satisfies SoS constraints on average”

where

degree-D polynomials in G

approximate SoS solution with $OBJ \geq k$ exists

≈ strong duality

no spectral algorithm

robustness + SDP duality

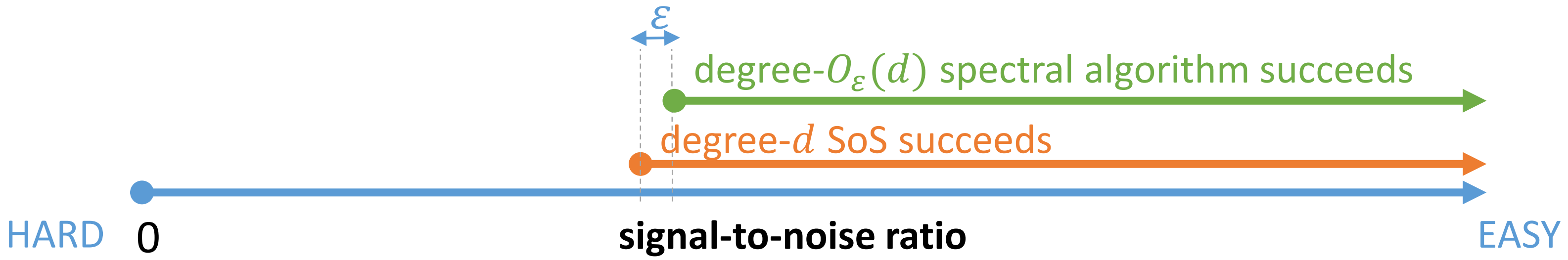
exact SoS solution with $OBJ \geq k$ exists

re-randomize coordinates

Concluding

Theorem

SoS and spectral algorithms are equivalent for average-case problems.



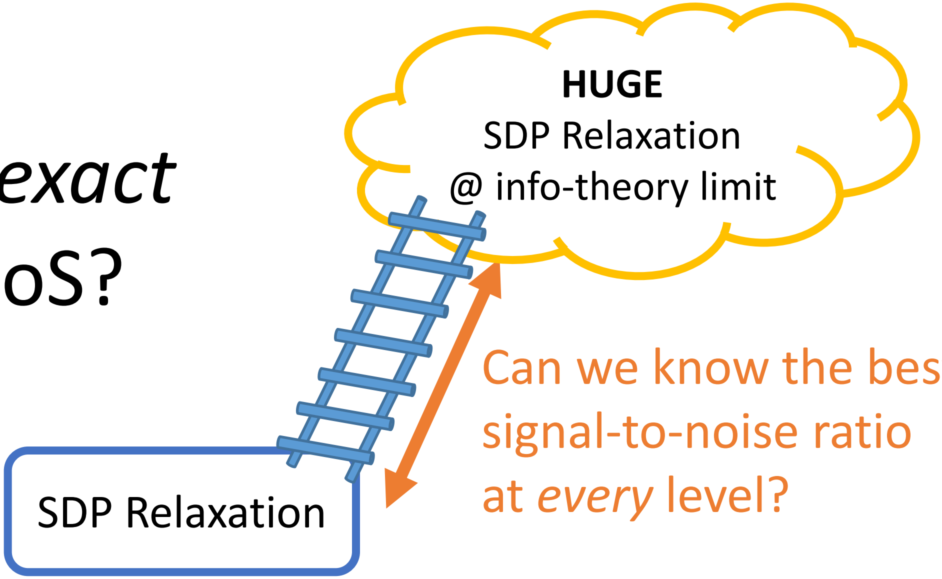
ε

degree-$O_\varepsilon(d)$ spectral algorithm succeeds

degree-d SoS succeeds

HARD 0 **signal-to-noise ratio** EASY

The future: can we characterize *exact* performance of average-case SoS?



HUGE
SDP Relaxation
@ info-theory limit

SDP Relaxation

Can we know the best signal-to-noise ratio at *every* level?

Thanks!