Random Optimisation Problems

Amin Coja-Oghlan

partly based on joint work with
Philipp Loick, Balazs Mezei, Greg Sorkin
thanks to Philipp for helping with the pictures!

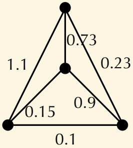
Goethe Universität Frankfurt

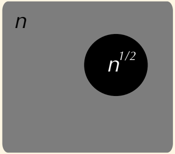
Part 1

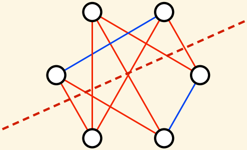
Motivation

- what are random optimisation problems?
- and what are they good for?

What are random optimisation problems?



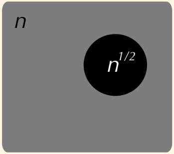


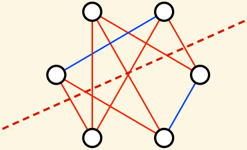


Examples

- minimum spanning trees with random weights [F85,...]
- planted clique [AKS98,...]
- MAX CUT on random graphs [CGHS04,...]

The average case myth

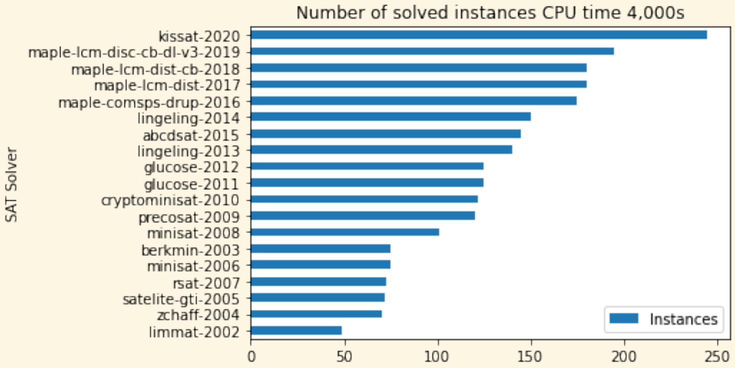




Hypothesis: hard instances are rare…

- …but real-world instances are easy!
- random models of real-world problems?
- semi-random, smoothed, power law… [FK01,ST01,FR19]
- where are the really hard problems?

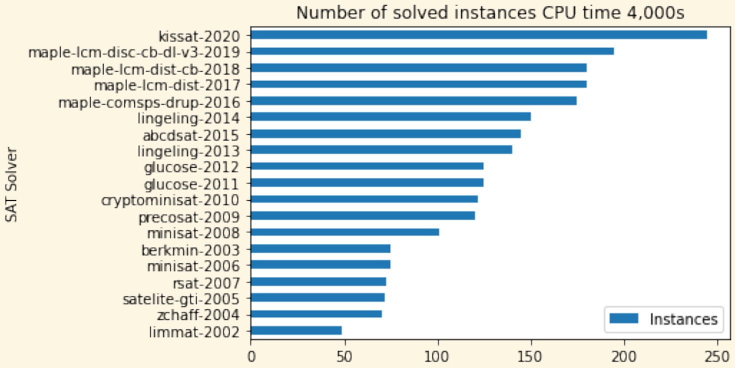
The average case myth



Example: SAT solving

- 10m variables, 32m clauses
- 2020 winner KISSAT based on clause learning [BFFH2020]
- real-world SAT instances solved on an industrial scale

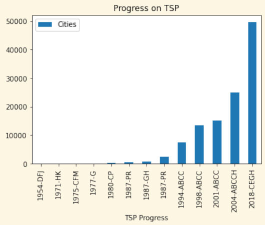
The average case myth

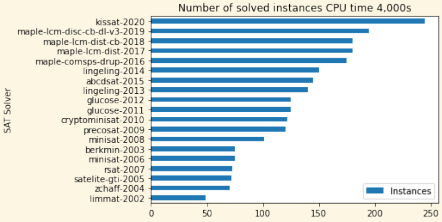


Example: random k-SAT

- satisfiability threshold $m/n \sim 2^k \log 2$ [DSS15]
- algorithmic barrier $m/n \sim 2^k \log(k)/k$ [ACO08,CO09]

The average case myth

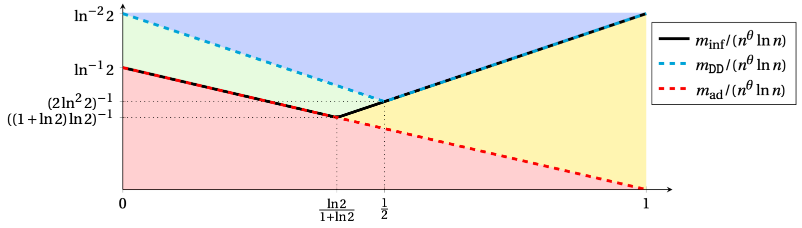




Hypothesis: hard instances are rare…

- …but real-world instances are easy! *[they are structured]*
- can random instances model real-world problems? *[not really]*
- semi-random, smoothed, power law… *[excess entropy]*
- where are the really hard problems? *[try random ones!]*

The probabilistic method



Success stories [personal selection]

- low-density parity check codes [G63,KRU11]
- compressed sensing [KMSSZ12,DJM13]
- group testing [COGHKL20]

A glimpse of complexity?

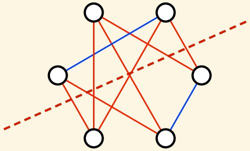
- proof complexity [BSW00,. . .]
- planted clique vs SOS [DM15,BHKKMP19,. . .]
- k-SAT refutation [F02,FO07,COGL04,KMODW17,. . .]
- overlap gap property [GS14,GZ19,BAWZ20,. . .]
- statistical algorithms [FPV13,SW20]

Part 2

Techniques

- what can we prove about random optimisation problems?
- what techniques do we have at our disposal?

Running example

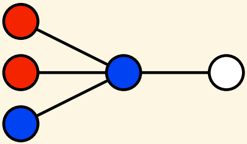


MaxCut on random regular graphs

$\mathbb{G} = \mathbb{G}(n, d)$ random d-regular graph of order n

$$\text{MAXCUT}(\mathbb{G}) = \max_{\sigma \in \{\pm 1\}^n} \sum_{vw \in E(\mathbb{G})} \frac{1 - \sigma_v \sigma_w}{2}.$$

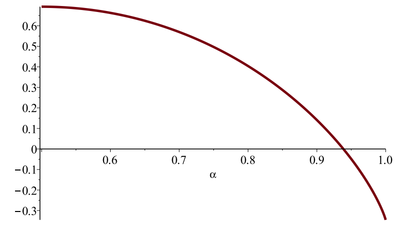
Combinatorial bounds



Greedy algorithms [DDSW03]

- assign each vertex the minority spin amongst its neighbours
- method of differential equations

Combinatorial bounds

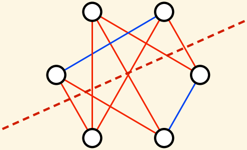


The first moment bound

The expected number of cuts of size $\alpha d n/2$ equals

$$\exp\left(n\left[\left(1-\frac{d}{2}\right)\log 2+\frac{d}{2}H(\alpha)+o(1)\right]\right)$$

The physics perspective



The Ising antiferromagnet

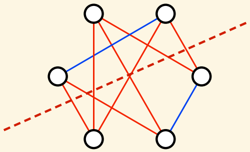
For an inverse temperature $\beta \geq 0$ introduce

$$\mathscr{H}_{\mathbb{G}}(\sigma) = \sum_{vw \in E(\mathbb{G})} \frac{1 + \sigma_v \sigma_w}{2}$$

$$Z_{\mathbb{G},\beta} = \sum_{\sigma \in \{\pm 1\}^n} \exp(-\beta \mathscr{H}_{\mathbb{G}}(\sigma))$$

$$\mu_{\mathbb{G},\beta}(\sigma) = \exp(-\beta \mathscr{H}_{\mathbb{G}}(\sigma)) / Z_{\mathbb{G},\beta}$$

The physics perspective



The Ising antiferromagnet

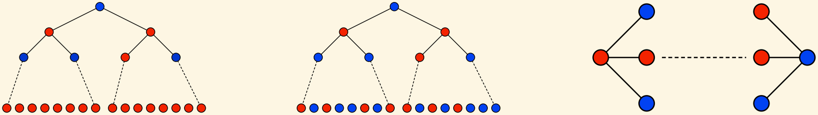
- The average number of edges cut equals

$$\frac{dn}{2} - \left\langle \mathcal{H}_{\mathbb{G}}(\boldsymbol{\sigma}), \mu_{\mathbb{G},\beta} \right\rangle = \frac{dn}{2} + \frac{\partial}{\partial \beta} \log Z_{\mathbb{G},\beta}$$

- Taking $\beta \to \infty$ yields

$$\text{MAXCUT}(\mathbb{G}) = \lim_{\beta \to \infty} \frac{dn}{2} + \frac{\partial}{\partial \beta} \log Z_{\mathbb{G},\beta}$$

Replica symmetry breaking



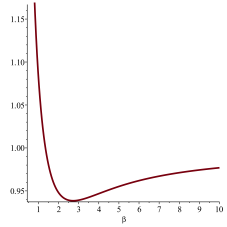
Absence of long-range correlations [KMRTSZ07]

Gibbs uniqueness: the root decouples from the boundary

Non-reconstruction: decoupling from a typical boundary

Static replica symmetry: $\mathbb{E}\left|\mu_{\mathbb{G},\beta}\left(\{\boldsymbol{\sigma}_{v_1} = \boldsymbol{\sigma}_{v_2}\}\right) - \frac{1}{2}\right| = o(1)$

Replica symmetry breaking



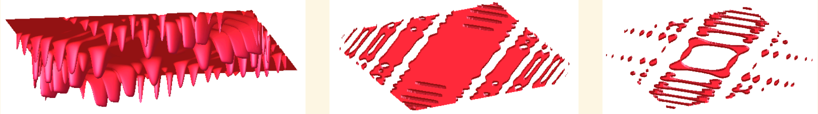
First moment redux

- assume static replica symmetry, assemble $\mathbb{G}$ one edge at a time
- every time we insert a new edge $\boldsymbol{e} = \boldsymbol{v}\boldsymbol{w}$,

$$\frac{Z_{\mathbb{G},\beta}(\mathbb{G}+\boldsymbol{e})}{Z_{\mathbb{G},\beta}} \sim 1-(1-\mathrm{e}^{-\beta})\mu_{\mathbb{G},\beta}(\{\boldsymbol{\sigma}_{\boldsymbol{v}}=\boldsymbol{\sigma}_{\boldsymbol{w}}\}) \sim \frac{1+\mathrm{e}^{-\beta}}{2}$$

$$\frac{1}{n}\log Z_{\mathbb{G},\beta} \sim \log 2 + \frac{d}{2}\log\frac{1+\mathrm{e}^{-\beta}}{2} = \frac{1}{n}\log \mathbb{E}[Z_{\mathbb{G},\beta}]$$

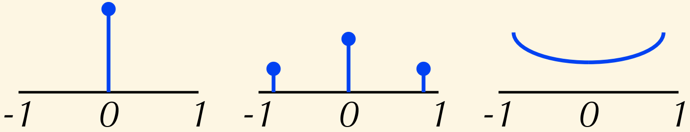
Replica symmetry breaking



Pure states [KMRTSZ07]

low β	the cross-section is contiguous
high β	decomposition into separate pure states

Replica symmetry breaking



The overlap [KMRTSZ07]

replica symmetry the overlap $\boldsymbol{\sigma} \cdot \boldsymbol{\sigma}' / n$ concentrates about 0

1-step replica symmetry breaking concentration on two points

full replica symmetry breaking no overlap concentration

The replica symmetry breaking transition

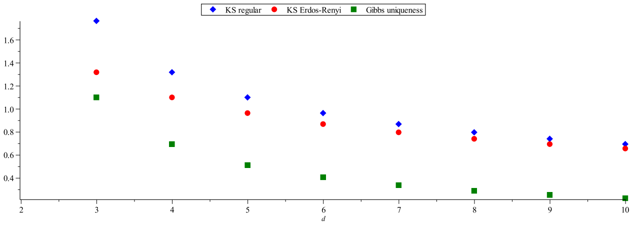
The free energy

The singularities of

$$\phi_d(\beta) = \lim_{n\to\infty} \frac{1}{n} \mathbb{E}[\log Z_{\mathbb{G},\beta}]$$

are called *phase transitions*.

The replica symmetry breaking transition



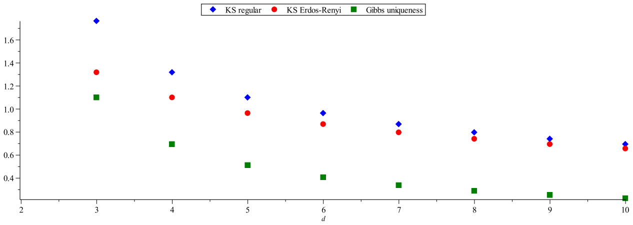
Theorem [COLMS20]

For any $d \geq 3$ we have

$$\phi_d(\beta) = \log 2 + \frac{d}{2} \log \frac{1 + \mathrm{e}^{-\beta}}{2} \qquad \text{if } \beta \leq \beta^*(d) = \log \frac{\sqrt{d-1}+1}{\sqrt{d-1}-1}$$

$$\phi_d(\beta) < \log 2 + \frac{d}{2} \log \frac{1 + \mathrm{e}^{-\beta}}{2} \qquad \text{if } \beta > \beta^*(d)$$

The replica symmetry breaking transition



Non-reconstruction/Kesten-Stigum

- broadcasting on the $(d-1)$-ary tree
- reconstruction threshold at $\beta^*(d) = \log \frac{\sqrt{d-1}+1}{\sqrt{d-1}-1}$
- Gibbs uniqueness threshold at $\beta^\dagger(d) = \log \frac{d}{d-2}$

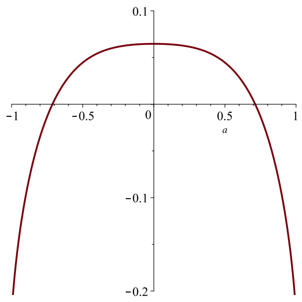
First attempt

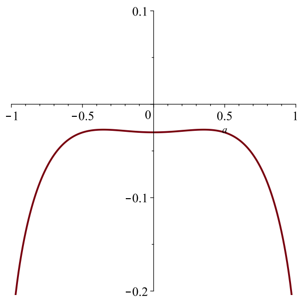
The second moment method

$$\begin{aligned}\mathbb{E}[Z_{\mathbb{G},\beta}^2] &= \sum_{\sigma,\sigma'\in\{\pm 1\}^n} \mathbb{E}\left[\exp(-\beta\mathcal{H}_{\mathbb{G}}(\sigma)-\beta\mathcal{H}_{\mathbb{G}}(\sigma'))\right] \\ &= \sum_{a=-n}^{n}\sum_{\sigma\cdot\sigma'=a} \mathbb{E}\left[\exp(-\beta\mathcal{H}_{\mathbb{G}}(\sigma)-\beta\mathcal{H}_{\mathbb{G}}(\sigma'))\right] \\ &= \exp\left(n\max_{\alpha\in[-1,1]} f_{d,\beta}(\alpha)+o(n)\right)\end{aligned}$$

$$f_{d,\beta}(\alpha) = (1-d)\log 2 + H\left(\frac{1+\alpha}{2}\right) + \frac{d}{2}\log\left((1+\mathrm{e}^{-\beta})^2+\alpha^2(1-\mathrm{e}^{-\beta})^2\right).$$

First attempt

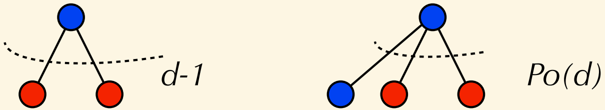




The second moment method

- $\mathbb{E}[Z^2_{\mathbb{G},\beta}] = O(\mathbb{E}[Z_{\mathbb{G},\beta}]^2)$ iff $f_{d,\beta}(\alpha)$ attains its max at $\alpha = 0$
- this is the case iff $\beta < \beta^*(d+1) < \beta^*(d)$

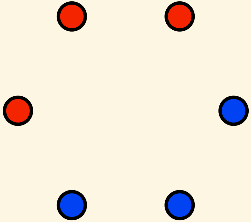
First attempt



The Erdős-Rényi model

- identical first/second moments
- Po(d) offspring in the Erdős-Rényi model
- phase transition at $\beta^*(d+1)$ [MNS15]

The planted model



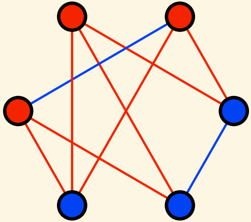
…aka stochastic block model

- draw a configuration $\boldsymbol{\sigma}^* \in \{\pm 1\}^n$
- draw a d-regular graph $\mathbb{G}^*$ from

$$\mathbb{P}\left[\mathbb{G}^* = G \mid \boldsymbol{\sigma}^*\right] \propto \exp(-\beta \mathscr{H}_G(\boldsymbol{\sigma}^*))$$

- $\mathbb{P}[\mathbb{G}^* \in \mathscr{E}] = \Theta(\mathbb{E}[Z_{\mathbb{G},\beta} \cdot \mathbb{1}\{\mathbb{G} \in \mathscr{E}\}])$

The planted model



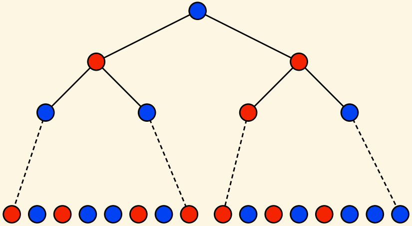
…aka stochastic block model

- draw a configuration $\boldsymbol{\sigma}^* \in \{\pm 1\}^n$
- draw a d-regular graph $\mathbb{G}^*$ from

$$\mathbb{P}\left[\mathbb{G}^* = G \mid \boldsymbol{\sigma}^*\right] \propto \exp(-\beta \mathscr{H}_G(\boldsymbol{\sigma}^*))$$

- $\mathbb{P}[\mathbb{G}^* \in \mathscr{E}] = \Theta(\mathbb{E}[Z_{\mathbb{G},\beta} \cdot \mathbb{1}\{\mathbb{G} \in \mathscr{E}\}])$

The planted model



The broadcasting process

- locally $(\mathbb{G}^*, \boldsymbol{\sigma}^*)$ resembles the broadcasing process
- $\Rightarrow \mu_{\mathbb{G}^*,\beta}$ has static replica symmetry for $\beta < \beta^*(d)$
- $\Rightarrow$ overlap concentrates about zero for $\beta < \beta^*(d)$ [cf. MNS15]

$$\mu_{\mathbb{G}^*,\beta}(\{\boldsymbol{\sigma} \cdot \boldsymbol{\sigma}'/n < \varepsilon\}) \sim 1$$

The planted model

The truncated second moment

- $\mathscr{E} = \left\{ \mu_{\mathbb{G},\beta}(\{\boldsymbol{\sigma} \cdot \boldsymbol{\sigma}' / n < \varepsilon\}) \sim 1 \right\}$
- $\mathbb{E}[Z_{\mathbb{G},\beta} \cdot \mathbb{1}\{\mathbb{G} \in \mathscr{E}\}] \sim \mathbb{E}[Z_{\mathbb{G},\beta}]$
- $\mathbb{E}[Z^2_{\mathbb{G},\beta} \cdot \mathbb{1}\{\mathbb{G} \in \mathscr{E}\}] = \mathbb{E}[Z_{\mathbb{G},\beta}]^2 \cdot \exp(\varepsilon^2 n)$
- Paley-Zygmund inequality

$$\Rightarrow \qquad Z_{\mathbb{G},\beta} \geq Z_{\mathbb{G},\beta} \cdot \mathbb{1}\{\mathbb{G} \in \mathscr{E}\} \geq \mathbb{E}[Z_{\mathbb{G},\beta}] \cdot \exp(o(n)) \text{ w.h.p.}$$

The first part of the theorem follows.

The planted model diverges

Quiet vs noisy planting

- $\log Z_{\mathbb{G},\beta}$ is tightly concetrated about $\mathbb{E}[\log Z_{\mathbb{G},\beta}]$
- $\Rightarrow$ if $\mathbb{E}[\log Z_{\mathbb{G},\beta}] \sim \log \mathbb{E}[Z_{\mathbb{G},\beta}]$, then $\mathbb{G} \approx \mathbb{G}^*$

$$\lim_{n\to\infty} \frac{1}{n}\mathbb{E}[\log Z_{\mathbb{G}^*,\beta}] > \log 2 + \frac{d}{2}\log\frac{1+\mathrm{e}^{-\beta}}{2} \geq \phi_d(\beta)$$

$$\Rightarrow \quad \phi_d(\beta) < \log 2 + \frac{d}{2}\log\frac{1+\mathrm{e}^{-\beta}}{2}$$

The planted model diverges

The planted free energy [COHKLMPP20]

$$\lim_{n\to\infty}\frac{1}{n}\mathbb{E}[\log Z_{\mathbb{G}^*,\beta}] = \sup_{\pi}\mathscr{B}^*_{d,\beta}(\pi)$$

$$\mathscr{B}^*_{d,\beta}(\pi) = \mathbb{E}\Bigg[\frac{\Lambda\left(\sum_{\sigma=\pm1}\prod_{i=1}^{d} 1-(1-\mathrm{e}^{-\beta})(1+\sigma\boldsymbol{\rho}_i)/2\right)}{2^{1-d}(1+\mathrm{e}^{-\beta})^d} - \frac{d\Lambda\left(1-(1-\mathrm{e}^{-\beta})(1+\boldsymbol{\rho}_1\boldsymbol{\rho}_2)/2\right)}{1+\mathrm{e}^{-\beta}}\Bigg]$$

$$\Lambda(x) = x\log x$$

The planted model diverges

Lemma

For $\beta > \beta^*(d)$ there is $\varepsilon > 0$ such that $\pi_\varepsilon = \frac{1}{2}(\delta_{-\varepsilon} + \delta_\varepsilon)$ satisfies

$$\mathscr{B}^*_{d,\beta}(\pi_\varepsilon) > \log 2 + \frac{d}{2} \log \frac{1 + \mathrm{e}^{-\beta}}{2}$$

Proof

- Compute the first four derivatives of $\varepsilon \mapsto \mathscr{B}^*_{d,\beta}(\pi_\varepsilon)$
- (seriously…)

We thus obtain the second part of the theorem.

The interpolation method

Theorem [SSZ16]

For any $d \geq 3, \beta > 0$, any probability measure r on $(-1,1)$ and any $0 < y < 1$ we have

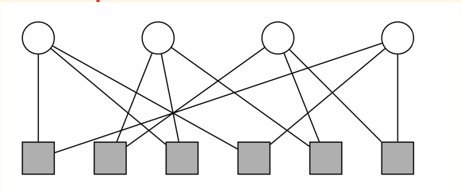
$$\phi_d(\beta) \leq \frac{1}{y} \log \mathbb{E}[\boldsymbol{X}^y] - \frac{d}{2y} \log \mathbb{E}[\boldsymbol{Y}^y]$$

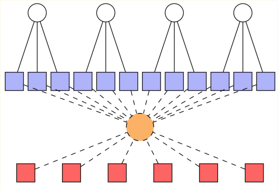
$$\boldsymbol{X} = \sum_{\tau=\pm 1} \prod_{i=1}^{d} 1 - (1 - \mathrm{e}^{-\beta}) \frac{1 + \tau \boldsymbol{\rho}_i}{2}$$

$$\boldsymbol{Y} = 1 - (1 - \mathrm{e}^{-\beta}) \sum_{\tau=\pm 1} \frac{(1 + \tau \boldsymbol{\rho}_1)(1 + \tau \boldsymbol{\rho}_2)}{4}$$

Regular variant of [G03, FL03, PT04].

The interpolation method





Pure states and Poisson–Dirichlet weights

- auxiliary model that represents the 1-step rsb scenario
- interpolation along time $t \in [0,1]$
- *Hypothesis:* the weights of the pure states are Poisson–Dirichlet
- to bound MAX CUT we use $\mathrm{MAXCUT}(\mathbb{G}) \leq \frac{dn}{2} + \frac{1}{\beta}\log Z_{\mathbb{G},\beta}$

The interpolation method

Theorem

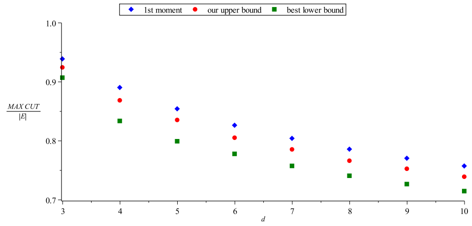
For any $d \geq 3$ we have

$$\mathrm{MAXCUT}(\mathbb{G}) \leq \frac{dn}{2}\left(1 + \inf_{\alpha,z} - \frac{\log \zeta \mathscr{A}^d \xi}{\log z} + \frac{d \log(1 - 2\alpha^2 + 2\alpha^2 z)}{2 \log z}\right)$$

$$\zeta = (1, 0, 0, \ldots) \qquad \xi = (1, z^{-1/2}, z^{-1}, \ldots)^T$$

$$\mathscr{A} = (1 - 2\alpha)\mathrm{id} + 2\alpha\sqrt{z}\begin{bmatrix} 0 & 1 & 0 & \cdots & \cdots & \cdots \\ \frac{1}{2} & 0 & \frac{1}{2} & \ddots & & \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & \ddots & \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots \\ \vdots & & \ddots & \ddots & \ddots & \ddots \\ \vdots & & & 0 & 1 & 0 \end{bmatrix}$$

The interpolation method



Some numbers

d	3	4	5	6	7	8	9	10
1st mmt	0.9385	0.8900	0.8539	0.8260	0.8038	0.7855	0.7701	0.7570
new bound	0.9241	0.8683	0.8350	0.8049	0.7851	0.7659	0.7523	0.7388
greedy	0.9067	0.8333	0.7989	0.7775	0.7571	0.7404	0.7263	0.7144

Match the numbers of [ZB10].

Outlook

- Bethe states and precise variational formulas [COP19]
- convergence to Sherrington–Kirkpatrick as $d \to \infty$ [DMS19]
- proof of the 1RSB formula in k-NAESAT [SSZ16]
- breaking of 1RSB [BSZ19]
- *open:* the Zdeborová–Boettcher conjecture [ZB10]
- *open:* better MAX CUT algorithms [see Eliran's talk]
- *open:* planted MAX CUT [CO05]
- *open:* ultrametricity and the Gibbs measure [P13]

Summary

Random optimisaton problems

- hard but rewarding
- algorithmic challenges and new algorithms
- probabilistic constructions
- techniques:
  - physics-enhanced moment methods
  - the planted model
  - spatial mixing and asymptotic Gibbs measures
  - the interpolation method
  - coupling arguments